



## (51) International Patent Classification:

G10L 21/0364 (2013.01) H03G 7/00 (2006.01)  
H03G 3/30 (2006.01) H03G 5/00 (2006.01)  
H03G 5/16 (2006.01) G10L 25/78 (2013.01)  
H03G 3/32 (2006.01) G10L 25/87 (2013.01)  
H03G 3/24 (2006.01)

## (21) International Application Number:

PCT/EP2018/069374

## (22) International Filing Date:

17 July 2018 (17.07.2018)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

17181799.2 18 July 2017 (18.07.2017) EP

(71) Applicant: **HARMAN BECKER AUTOMOTIVE SYSTEMS GMBH** [DE/DE]; Becker-Goering-Str. 16, 76307 Karlsbad (DE).

(72) Inventor: **CHRISTOPH, Markus**; Am Gstütt 4, 94315 Straubing (DE).

(74) Agent: **WESTPHAL, MUSSGUG & PARTNER MBB**; Werinherstr. 79, 81541 München (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

## (54) Title: SPEECH SIGNAL LEVELING

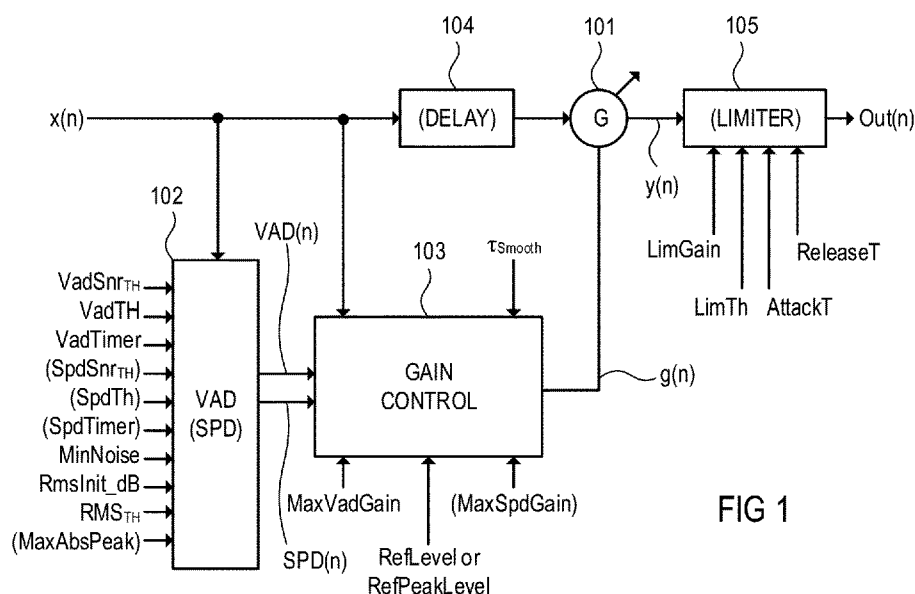


FIG 1

(57) Abstract: A speech signal leveling system and method include generating an output signal by applying a frequency-dependent or frequency-independent controllable gain to an input signal, the gain being dependent on a gain control signal, and generating at least one speech detection signal indicative of voice components contained in the input signal. The system and method further include generating the gain control signal based on the input signal and the at least one speech detection signal, controlling the controllable-gain block to amplify or attenuate the input signal to have a predetermined mean or maximum or absolute peak signal level as long as voice components are detected in the input signal.

**Published:**

— *with international search report (Art. 21(3))*

## SPEECH SIGNAL LEVELING

## BACKGROUND

5    **1. Technical Field**

[0001]     The disclosure relates to a system and method (generally referred to as a “system”) for speech signal leveling.

**2. Related Art**

10    [0002]     In speech signal processing such as speech recognition and hands-free communication, a well leveled output speech signal is crucial for an adequate recognition rate of a speech recognition system or an adequate intelligibility at a far-end side of a hands-free system. In a simple approach, an automatic gain control (AGC) may be based on a (peak) limiter whose (input) gain is set in such a way that speech, even when softly spoken or spoken from a far distance, is amplified to still deliver a well-leveled output  
15    signal, i.e. a speech signal whose peaks are ideally full scale of an acceptable maximum amplitude range. On the other hand, any speech spoken at closer distance with a loud voice is attenuated to full scale by the limiter. In this way, the signal-to-noise ratio (SNR) will decrease if the talker is at a shorter distance to an input microphone as the limiter will be more often active so that undesired noise and/or residual echoes are not limited in  
20    contrast to the speech which exceeds full scale and is limited to full scale by the (peak) limiter. If, however, a speaker is too far away from the microphone or speaks softly, the speech output signal will not be full scale, which again may decrease the SNR. As a consequence, a speech signal leveler is desired which automatically adjusts the speech level to such dynamically changing situations.

## 25    SUMMARY

[0003]     A speech signal leveling system includes a controllable-gain block configured to receive an input signal and to generate an output signal by applying a frequency-dependent or frequency-independent controllable gain to the input signal, the

gain being dependent on a gain control signal. The system further includes a speech detecting block configured to receive the input signal and to generate at least one speech detection signal indicative of voice components contained in the input signal. The system further includes a gain control block configured to receive the input signal and the at least one speech detection signal, and to generate the gain control signal based on the input signal and the at least one speech detection signal. The gain control block is further configured to control the controllable-gain block to amplify or attenuate the input signal to have a predetermined mean or maximum or absolute peak signal level as long as voice components are detected in the input signal.

[0004] A speech signal leveling method includes generating an output signal by applying a frequency-dependent or frequency-independent controllable gain to an input signal, the gain being dependent on a gain control signal, and generating at least one speech detection signal indicative of voice components contained in the input signal. The method further includes generating the gain control signal based on the input signal and the at least one speech detection signal, and controlling the controllable-gain block to amplify or attenuate the input signal to have a predetermined mean or maximum or absolute peak signal level as long as voice components are detected in the input signal.

[0005] Other systems, methods, features and advantages will be, or will become, apparent to one with skill in the art upon examination of the following detailed description and appended figures. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the invention, and be protected by the following claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The system may be better understood with reference to the following drawings and description. The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention. Moreover, in the figures, like referenced numerals designate corresponding parts throughout the different views.

[0007] Figure 1 is a simplified schematic diagram illustrating a signal processing structure and a signal flow of an exemplary speech leveling system.

[0008] Figure 2 is a detailed schematic diagram illustrating a signal processing structure and signal flow of an exemplary implementation of the speech leveling system shown in Figure 1.

[0009] Figure 3 is a flow chart illustrating an exemplary speech leveling method.

5 [0010] Figure 4 is a simplified schematic diagram illustrating a signal processing structure and signal flow of another exemplary speech leveling system.

[0011] Figure 5 is a detailed schematic diagram illustrating a signal processing structure and signal flow of an exemplary implementation of the speech leveling system shown in Figure 4.

10 [0012] Figure 6 is a detailed schematic diagram illustrating a signal processing structure and signal flow of a modification of the speech leveling system shown in Figure 4.

[0013] Figure 7 is a detailed schematic diagram illustrating a signal processing structure and signal flow of another modification of the speech leveling system shown in  
15 Figure 4.

#### DETAILED DESCRIPTION

[0014] In the following description of exemplary speech leveling systems and methods, a block is understood to be a hardware system or an element thereof with at least one of: a processing unit (such as a controller, processor etc.) executing software  
20 and a dedicated circuit structure for implementing a desired signal transferring or processing function. Blocks may be connected to or in another way operatively coupled with each other to provide a specific signal flow structure and to generate specific signals as detailed below. A characteristic of speech (voice) is the presence of silent intervals (empty pauses) and vocalizations (filled pauses) that do not have a lexical meaning. An  
25 empty and a filled pause are likely to coincide with boundaries, realized as a silent interval of varying length, at clause and paragraph level and often mark the boundaries of narrative units. Pauses in speech are typically a multi-determined phenomenon attributable to physical, socio-psychological, communicative, linguistic and cognitive causes.

[0015] Figure 1 is a schematic diagram illustrating the signal processing structure and the signal flow of an exemplary speech leveling system (which may perform an exemplary speech leveling method). The speech leveling system shown in Figure 1 includes a controllable-gain block 101 (e.g., a controllable amplifier, controllable attenuator, controllable filter, multiplier, etc.) that receives an input signal  $x(n)$  and provides an output signal  $y(n)$ . The output signal  $y(n)$  is the input signal  $x(n)$  to which a frequency-dependent or frequency-independent controllable gain  $G$  has been applied. The controllable-gain block 101, i.e., its gain  $G$ , is controlled by way of a gain control signal  $g(n)$ . The input signal  $x(n)$  is also supplied to a speech detecting block 102 that generates at least one speech detection signal indicative of voice components contained in the input signal  $x(n)$ . In the example system shown, two speech detection signals, a voice activity detection signal  $VAD(n)$ , e.g., a VAD flag, and a speech pause detection signal  $SPD(n)$ , e.g., an SPD flag, are provided whose generation and characteristics are detailed further below.

[0016] The input signal  $x(n)$  is also supplied to a gain control block 103 which additionally receives the speech detection signals and generates the gain control signal  $g(n)$  based on the input signal  $x(n)$  and the speech detection signal(s), e.g., the voice activity detection signal  $VAD(n)$  and optionally the speech pause detection signal  $SPD(n)$ . The gain control block 103 controls the controllable-gain block 101 so that the input signal  $x(n)$  is amplified or attenuated to have a predetermined mean or maximum or absolute peak signal level as long as no voice components are detected in the input signal  $x(n)$ . For example, the input signal  $x(n)$  may be blocked, i.e., the output signal  $y(n)$  is (almost) zero, as long as no voice components are detected in the input signal  $x(n)$ .

[0017] The speech leveling system shown in Figure 1 may further include an optional delay block 104 that delays by a certain time (e.g., a certain number of frames or samples) the input signal  $x(n)$  before it is supplied to the controllable-gain block 101 such that the input signal  $x(n)$  supplied to the controllable-gain block 101 is delayed with respect to the input signal  $x(n)$  as it is supplied to the speech detecting block 102 and the gain control block 103. Situations in which the delay block 104 may be utilized are outlined further below.

[0018] The speech detecting block 102 may further receive, e.g., from a register or memory (both not shown), an SNR threshold  $VadSnr_{TH}$ , a threshold  $Vad_{TH}$ , and, a timer value  $Vad_{Timer}$ , for the voice activity detection, and, optionally, a SNR threshold  $SpdSnr_{TH}$ , a threshold  $Spd_{TH}$ , and a timer value  $Spd_{Timer}$  for the optional speech pause detection.

5 Further, at least one of a minimum noise threshold  $MinNoise$ , an initial root mean square (RMS) value  $RmsInit\_dB$ , a root mean square threshold  $RMS_{TH}$ , and a threshold  $MaxAbsPeak$  may be provided to the speech detecting block 102. The gain control block 103 may further receive, e.g., from a register or memory (both not shown), an optional smoothing time parameter  $\tau_{Smooth}$ , at least one of tuning parameter  $MaxVadGain$  and  
 10 tuning parameter  $MaxSpdGain$ , and a reference level  $RefLevel$  or a reference peak level  $RefPeakLevel$ . Optionally, a (peak) limiter block 105 may operatively be connected downstream of the controllable-gain block 101 to receive the output signal  $y(n)$  and to provide a limited output signal  $Out(n)$ . The (peak) limiter block 105 may further receive a predetermined limiter gain  $LimGain$ , a limiter threshold  $LimTh$ , an attack time  $AttackT$ ,  
 15 and a release time  $ReleaseT$ .

[0019] Figure 2 is a schematic diagram illustrating the signal processing structure and the signal flow of another exemplary speech leveling system (which performs another exemplary speech leveling method). In the speech leveling system shown in Figure 2, a  
 20 speech detector block 201 may include transforming the input signal  $x(n)$  from the time domain into the frequency domain, which may be performed by a time-frequency transformation block 202. The spectral components of the input signal  $x(n)$  can be obtained by a variety of ways, including band pass filtering and Fourier transformation. In one approach, a Discrete Fourier Transform (DFT) or Fast Fourier Transform (FFT) may be used to transform sequential blocks of  $N$  points of the input signal. A window  
 25 function, such as a Hanning window, may be applied, in which case an overlap of  $N/2$  points can be used. The DFT can be used at each frequency bin in the input signal. Alternatively, the FFT can be used over the whole frequency band occupied by the input signal. The spectrum is stored for each frequency bin within the input signal band.

[0020] In the present example, time-frequency transformation block 202 applies an  
 30 FFT with optional windowing (not shown) to the input signal  $x(n)$  in the time domain to

generate a signal  $X(\omega)$  in the frequency domain. The signal  $X(\omega)$  may be optionally smoothed utilizing, e.g., a moving average filter of appropriate length.

[0021] The (optional) spectral smoothing may account for a plurality of frequency bins, which reduces the spectral resolution. The output of the time-frequency transformation block 202 is further smoothed by a multiplicity (M) of temporal smoothing blocks 203 that are operatively connected downstream of the time-frequency transformation block 202. The multiplicity of temporal smoothing blocks 203 combines frequency bin values over time to reduce the temporal dynamics in the output signal of time-frequency transformation block 202.

[0022] The multiplicity of temporal smoothing blocks 203 outputs temporal smoothed signals that may still contain impulsive distortions as well as background noise. A multiplicity (M) of noise estimation blocks 204 is operatively connected downstream of the multiplicity of temporal smoothing blocks 203 to smear out impulsive distortions such as speech in the outputs of the multiplicity of temporal smoothing blocks 203 and to estimate the current background noise. In order to reduce or avoid smearing of a desired signal such as music or voice signals, non-linear smoothing (not shown) may be employed in the multiplicity of noise estimation blocks 204.

[0023] Based on the smoothed signals from the multiplicity of temporal smoothing blocks 203 and the estimated quasi stationary background noise signals from the multiplicity of noise estimation blocks 204, the variations in the SNR can be determined (calculated), e.g., as frequency distribution of SNR values. Using variations in the SNR, noise can be differentiated from a desired speech or music signal. For example, a low SNR value may be linked to a variety of noise sources such as an air-conditioner, fan, an open window, or an electrical device such as a computer etc. The SNR may be evaluated in the time domain or in the frequency domain or in the sub-band domain.

[0024] In a multiplicity (M) of comparator blocks 206, output SNR values determined by the multiplicity (M) of SNR determination blocks 205 are compared to a pre-determined threshold, e.g., an SNR threshold  $VadSnr_{TH}$ . If the determined current SNR values are greater than the pre-determined threshold  $VadSnr_{TH}$ , flags  $VadSnrFlag$  will be set to a first logic value, e.g., '1'. Otherwise, the flags  $VadSnrFlag$  will be set to a

second logic value, e.g., '0'. The flags from the multiplicity of comparator blocks 206 are passed on to a summation block 207. The summation block 207 sums up the spectral flags  $VadSnrFlag$  from the multiplicity of comparator blocks 206 and outputs a time varying sum signal  $S_{VAD}(n)$ . Further, an RMS determination block 211 may provide a mean signal

5  $RMS(n)$  indicative of the (root mean square) mean of the input signal  $x(n)$ . The sum signal  $S_{VAD}(n)$  and the mean signal  $RMS(n)$  are passed to a comparator block 208 where the sum signal  $S_{VAD}(n)$  is compared to a pre-determined sum threshold  $VAD_{TH}$ , and the mean signal  $RMS(n)$  is compared to a pre-determined threshold  $RMS_{TH}$ . If the current value of the sum signal  $S_{VAD}(n)$  is greater than the pre-determined sum threshold  $VAD_{TH}$ , and the

10 mean signal  $RMS(n)$  is greater than the pre-determined threshold  $RMS_{TH}$ , a flag  $S'_{VAD}(n)$ , which can be considered an updated sum signal  $S_{VAD}(n)$ , will be set to a first logic state, e.g., '1' (one). Otherwise the flag  $S'_{VAD}(n)$  will be set to another logical state, e.g., '0' (zero). Instead of the RMS determination block 211, a maximum absolute peak determination block (not shown) that is configured to determine a maximum absolute

15 peak value of the input signal  $x(n)$  may be used to subsequently compare the maximum absolute peak value to a maximum absolute peak threshold.

**[0025]** The output signal of the comparator block 208, i.e., flag  $S'_{VAD}(n)$ , is passed to a voice activity detection (VAD) timer block 209 that detects whether flag  $S'_{VAD}(n)$  is '0' or not. At a time  $n$ , if the flag  $S_{VAD}(n)$  is '0', the VAD timer block 209 sets a count

20  $T_{VAD}(n)$  to  $T_{VAD}(n-1)-1$ , i.e., decrements the previous count  $T_{VAD}(n-1)$  by one. If the flag  $S_{VAD}(n)$  is not '0', i.e., is '1', the count  $T_{VAD}(n)$  will be reset to an initialization value  $VadTimerInit$ . The output of the VAD timer block 209, i.e., count  $T_{VAD}(n)$ , is passed on to a voice activity detection (VAD) block 210 where the count  $T_{VAD}(n)$  is compared with a pre-determined threshold, e.g., zero. If the current count  $T_{VAD}(n)$  is greater than the pre-

25 determined threshold (zero), a flag  $VAD(n)$  will be set to '1', otherwise to '0'.

**[0026]** Additionally or alternatively, in a multiplicity ( $M$ ) of comparator blocks 212, the SNR values determined by the multiplicity ( $M$ ) of SNR determination blocks 205 are compared with a pre-determined threshold, e.g., an SNR threshold  $SpdSnr_{TH}$ . If the determined current SNR values are greater than the pre-determined threshold  $SpdSnr_{TH}$ ,

30 flags  $SpdSnrFlag$  will be set to, e.g., '1'. Otherwise, the flags  $SpdSnrFlag$  will be set to '0'. The flags from the multiplicity of comparator blocks 212 are passed on to a

summation block 213. The summation block 213 sums up the flags  $\text{SpdSnrFlag}$  from the multiplicity of comparator blocks 212 and outputs a time varying sum signal  $S_{\text{SPD}}(n)$ . The sum signal  $S_{\text{SPD}}(n)$  and the mean signal  $\text{RMS}(n)$  from RMS determination block 211 are passed to a comparator block 214 where the sum signal  $S_{\text{SPD}}(n)$  is compared to a pre-determined sum threshold  $\text{SPD}_{\text{TH}}$ , and the mean signal  $\text{RMS}(n)$  is compared to a pre-determined threshold  $\text{RMS}_{\text{TH}}$ . If the current value of the sum signal  $S_{\text{SPD}}(n)$  is greater than the pre-determined sum threshold  $\text{SPD}_{\text{TH}}$ , and the mean signal  $\text{RMS}(n)$  is greater than the pre-determined threshold  $\text{RMS}_{\text{TH}}$ , a flag  $S'_{\text{SPD}}(n)$ , which can be considered as updated sum signal  $S'_{\text{SPD}}(n)$ , will be set to, e.g., '1'. Otherwise the flag  $S'_{\text{SPD}}(n)$  will be set to '0'.

- 10 [0027] The output signal of the comparator block 214, i.e., flag  $S'_{\text{SPD}}(n)$ , is passed to a speech pause detection (SPD) timer block 215 that detects whether flag  $S'_{\text{SPD}}(n)$  is '0' or not. At a time  $n$ , if the flag  $S'_{\text{SPD}}(n)$  is '0', the SPD timer block 215 sets a count  $T_{\text{SPD}}(n)$  to  $T_{\text{SPD}}(n-1)-1$ , i.e., decrements the previous count  $T_{\text{SPD}}(n-1)$  by one. If the flag  $S'_{\text{SPD}}(n)$  is not '0', i.e., is '1', the count  $T_{\text{SPD}}(n)$  will be reset to an initialization value  $\text{SpdTimerInit}$ .
- 15 The output of the SPD timer block 215, i.e., count  $T_{\text{SPD}}(n)$ , is passed on to a speech pause detection (SPD) block 216 where the count  $T_{\text{SPD}}(n)$  is compared to a pre-determined threshold, e.g., zero. If the current count  $T_{\text{SPD}}(n)$  is smaller or equal than the pre-determined threshold (zero), a flag  $\text{SPD}(n)$  will be set to '1', otherwise to '0'.

- [0028] If flag  $\text{VAD}(n)$  is '1' at time  $n$ , voice activity is detected. If it is '0', no speech activity is detected. If flag  $\text{SPD}(n)$  is '1' at a time  $n$ , a speech pause is detected. If it is '0', no speech pause is detected. For example, initialization values  $\text{VadTimerInit}$  as utilized in VAD timer block 209 and  $\text{SpdTimerInit}$  as utilized in SPD timer block 215 may be different. SNR thresholds  $\text{VadSnr}_{\text{TH}}$  as utilized in comparator block 206 and  $\text{SpdSnr}_{\text{TH}}$  as utilized in comparator block 212 may have different values, for example. Sum thresholds  $\text{VAD}_{\text{TH}}$  and  $\text{SPD}_{\text{TH}}$  as utilized in comparator blocks 208 and 214 may be equal, for example.

- [0029] In the speech leveling system shown in Figure 2, a gain control block 217 includes a speech RMS level determination block 218 which receives the input signal  $x(n)$ , the flag  $\text{VAD}(n)$  and a previous sample of smoothed RMS signal  $\text{SRMS}(n-1)$ , and which outputs a signal  $\text{In}_{\text{RMS}}(n)$ . The signal  $\text{In}_{\text{RMS}}(n)$  corresponds to the input signal  $x(n)$  if the flag  $\text{VAD}(n)$  is '1', and is set to the value of the previous sample of the smoothed

RMS signal  $SRMS(n-1)$  otherwise. The signal  $In_{RMS}(n)$  is subject to a subsequent RMS determination (calculation) optionally utilizing the initial root mean square (RMS) value  $RmsInit\_dB$ . The value out of the RMS determination is supplied to a smoothing block 219 which provides a smoothed RMS signal  $SRMS(n)$ . The smoothed RMS signal  $SRMS(n)$  is, on the one hand, delayed by one sample to generate the previous sample of the smoothed RMS signal  $SRMS(n-1)$  which is fed back to the speech RMS level determination block 218, and, on the other hand, is supplied to a gain determination block 220 where a gain control signal  $Gain(n)$  is generated according to the ratio between a reference level  $RefLevel$  and the smoothed RMS signal  $SRMS(n)$ .

10 [0030] In an optional gain limitation block 221, the gain control signal  $Gain(n)$  is compared to a reference value  $MaxVadGain$ . As a consequence, if the gain control signal  $Gain(n)$  is greater than the reference value  $MaxVadGain$ , the output signal of the gain limitation block 221, limited gain control signal  $Gain'(n)$  which can be considered an updated gain control signal  $Gain(n)$ , is set to the reference voice activity value  $MaxVadGain$ , and otherwise to the gain control signal  $Gain(n)$ . In the present example, the limited gain control signal  $Gain'(n)$  is supplied to an SPD gain limitation block 222 which further receives the flags  $VAD(n)$  and  $SPD(n)$  as well as a reference speech pause value  $MaxSpdGain$ . If flag  $SPD(n)$  is '1' and flag  $VAD(n)$  is '0' at time  $n$ , a final gain control signal  $Gain''(n)$ , which can be considered as updated gain control signal  $Gain'(n)$ , is set to the reference speech pause value  $MaxSpdGain$ , and otherwise to the limited gain control signal  $Gain'(n)$ . Alternatively, block 222 may receive only the SPD flag. The final gain control signal  $Gain''(n)$  may optionally be smoothed over time by way of a smoothing filter block 226 before it controls a controllable gain block 223 which is supplied via an optional delay block 242 with the input signal  $x(n)$  and which outputs via an optional peak limiter 225 the limited output signal  $Out(n)$ .

[0031] In the system shown in Figure 2, two different tuning sets are used for the determination of the voice activity detection flag  $VAD(n)$  and the optional speech pause detection flag  $SPD(n)$ . However, both are based on the same spectral SNR values. As already mentioned above, for the determination of the voice activity detection flag  $VAD(n)$ , a higher value for the SNR threshold  $VadSnr_{TH}$  may be used than for the SNR threshold  $SpdSnr_{TH}$  for the determination the speech pause detection flag  $SPD(n)$ . This

allows, on the one hand, to only set the voice activity detection flag  $VAD(n)$  if it is sure that speech is present, hence the application of a high threshold, and, on the other hand, to set the speech pause detection flag  $SPD(n)$  to '1' only if it is sure that a speech pause occurs. This may be achieved by setting the respective SNR threshold to a lower value  
 5 such that even vague speech parts will result in a detected speech activity and, thus, avoid an erroneous setting of the speech pause detection flag  $SPD(n)$ .

[0032] Further, the numbers of spectral lines that exceed the given signal-to-noise ratio thresholds  $VadSnr_{TH}$  and  $SpdSnr_{TH}$  are determined (counted) and these numbers are compared to the thresholds  $VAD_{TH}$  for the voice activity detection part and  $SPD_{TH}$  for the  
 10 speech pause detection part. In parallel, also the RMS value of the input signal  $x(n)$ , or a signal corresponding to the input signal  $x(n)$  such as the output signal  $y(n)$  is calculated per frame whose size may depend on the length of an involved FFT and a chosen overlap. For example, the FFT length may be 512 taps, at a sampling frequency  $f_s$  of 16 kHz and the overlap may be set to 50 %, which leads to a frame size of 256 samples. This means  
 15 that after every block of samples (e.g., every 256th sample ) the RMS value of the input signal  $x(n)$ , or a signal corresponding to the input signal  $x(n)$  such as the output signal  $y(n)$ , is calculated from all block samples (e.g., 256 previous samples) and then compared to the RMS threshold  $RMS_{TH}$ . Only if the current RMS value  $RMS(n)$  exceeds the threshold  $RMS_{TH}$ , a positive decision of a previous stage that compares the numbers of  
 20 samples of spectral signal-to-noise ratios exceeding the thresholds  $VadSnr_{TH}$  and  $SpdSnr_{TH}$ , will be accepted (or not). If the current RMS value  $RMS(n)$  does not exceed the RMS threshold  $RMS_{TH}$ , neither a positive voice activity nor a positive speech pause decision will be made and the current status will be maintained instead.

[0033] The RMS processing part serves to avoid erroneous settings of the detection  
 25 flags for voice activity detection and speech pause detection in occasions where the signal-to-noise ratio is high but the level of the input signal is low. Since the signal-to-noise ratio is a relative value it is possible that high signal-to-noise ratios will occur, even if the level of the input signal is very low as, for example, in echo-only scenarios. In such a case, the level of the residual echo can be very low but its signal-to-noise ratio can still  
 30 be high and as such could erroneously trigger the flag for voice activity detection  $VAD(n)$  and the flag for speech pause detection  $SPD(n)$  if taken as sole decision-maker. To avoid

this, additionally at least one other absolute value may be employed before a final decision is made. One approach is to take the energy or the RMS value of the input signal  $x(n)$  or output signal  $y(n)$  into consideration. For example, the signal-to-noise ratio will only be considered if, e.g., the current RMS value exceeds a certain limit so that erroneous settings of the flag for voice activity detection  $VAD(n)$  and the flag for speech pause detection  $SPD(n)$  by very small input signals is avoided. In another approach, the minimum noise threshold  $MinNoise$  is set to a much higher value. Thus, very small input signals can be limited to this threshold and as a consequence the achievable signal-to-noise ratios will decrease in case of very small input signals.

**[0034]** In order to avoid that the gain rises to undesired high values, the residual echo may be amplified more or less to full scale and/or the gain may be limited to a maximum, e.g., by way of the optional tuning parameter  $MaxVadGain$ . Thereby, different timer values for voice activity detection and speech pause detection may be employed. Whereas the timer for the voice activity detection may be set to an average length of a word, the timer for the speech pause detection may be set to a higher value, to ensure that an utterance is securely finished before the time varying speech pause detection flag  $SPD(n)$  is set to '1'. In contrast, the time varying voice activity detection flag  $VAD(n)$  will be set to '0', if its initial timer value  $VadTimerInit$  has expired.

**[0035]** If the time varying VAD flag is set to '1', i.e. only in active speech sequences, the RMS value or, alternatively, the maximum absolute peak value of the input signal  $x(n)$  will be determined. Otherwise, the previous sample of the smoothed speech RMS value  $SRMS(n-1)$  will be passed on to the smoothing filter. In order to avoid undesired transient effects, RMS determination may optionally be initialized with a properly chosen initialization value  $RmsInit\_dB$ . As a smoothing filter, for example, an Infinite Impulse Response (IIR) low-pass filter with a relatively long smoothing time constant  $\tau$ , may be used. The current gain value  $Gain(n)$  is determined (calculated) by taking the ratio of the desired reference RMS value of the speech output signal, i.e., reference level  $RefLevel$  or alternatively reference peak level  $RefPeakLevel$ , and the output signal of this smoothing filter, i.e.,  $SRMS(n)$ , so that  $Gain(n) = RefLevel/SRMS(n)$ .

**[0036]** Further, an optional limitation of the maximum of the gain value  $Gain(n)$  may be applied, in order to avoid successive amplifications of very small input signals, as

already described above. Still another limitation of the gain value may optionally be applied, which may act as a level compensator in hands-free applications and which may be controlled by the time varying voice activity detection flag  $VAD(n)$  and/or the time varying speech pause detection flag  $SPD(n)$ . Thereby the gain value  $Gain(n)$  will be reduced to the given limit  $MaxSpdGain$  whenever a speech pause and no voice activity is detected, e.g. whenever the speech pause detection flag  $SPD(n)$  assumes '1' and the time varying voice activity detection flag  $VAD(n)$  assumes '0'. For example, no gain or even gain reduction will be applied to the input signal  $x(n)$ , if  $SPD(n) = '1'$  and  $VAD(n) = '0'$ .

[0037] The procedure outlined above, which can also be considered as "gating of the gain value", allows, e.g., to fulfill the requirements for a Terminal Coupling Loss weighted (TCLw) value in hands-free mode so that it is possible to meet the given requirement for a TCLw value of e.g.  $\geq -52$  dB, as the measurement of the TCLw value requires that this value has to be reached between an input signal, which is taken by a reference microphone in proximity to the microphones installed at the device and the final, electrical output signal of the device, while the device is playing at a defined playback volume, to reach a sound pressure level of  $L \approx -76$  dB SPL(A) in a distance of  $d = 1$  m from speaker(s) of the device. In hands-free mode it is also specified to deliver a minimum output speech RMS signal of about -24 dBFS, just as is the case in speech recognition mode. Without the before mentioned gating functionality of the gain determination unit, these somewhat contradicting requirements can be hardly fulfilled.

[0038] With the described system it is possible to securely sense speech pauses and as such to apply the needed gating to the gain value  $Gain(n)$ . A major challenge is to leave this gating stage quick enough once speech (from the near end talker) is active again. If this is not guaranteed, it may happen that the first parts of the on-setting speech may not be amplified as desired and as such cannot be understood, neither by speech recognition systems, nor by a human listener (far-end talker). One way to avoid this is to use a quickly reacting voice activity detection and/or speech pause detection. This can be realized by using a non-smoothed, instantaneous RMS value which will be compared to the given threshold as part of the decision. If this is not sufficient, the input signal can optionally be delayed by one or several frames, before the current gain  $Gain(n)$  is applied to it. This can be done as long as the maximally given latency is not exceeded.

[0039] Figure 3 is a flow chart illustrating an exemplary speech leveling method. The method includes a procedure 301 in which an output signal is generated by applying a frequency-dependent or frequency-independent controllable gain to an input signal, the gain being dependent on a gain control signal. In a procedure 302, at least one speech detection signal indicative of voice components contained in the input signal is generated. The method further includes a procedure 303 in which the gain control signal is generated based on the input signal and the at least one speech detection signal and the controllable-gain block is controlled to amplify or attenuate the input signal to have a predetermined mean signal level as long as no voice components are detected in the input signal, e.g., to block the input signal so that no output signal is provided as long as no voice components are detected in the input signal.

[0040] Figure 4 is a simplified schematic diagram illustrating a signal processing structure and signal flow of another exemplary speech leveling system. The system shown in Figure 4 includes a controllable-gain block 401 (e.g., a controllable amplifier, controllable attenuator, controllable filter, multiplier, etc.) that receives, as an input signal, an output signal  $\text{OutTfe}(n)$  sent to a far end and provides an output signal  $\text{Out}(n)$ . The output signal  $\text{Out}(n)$  is here the output signal  $\text{OutTfe}(n)$  to which a frequency-dependent or frequency-independent controllable gain  $G$  has been applied. The controllable-gain block 401, i.e., its gain  $G$ , is controlled by way of a gain control signal  $g(n)$ . The output signal  $\text{OutTfe}(n)$  is also supplied to a speech detecting block 402 that generates at least one speech detection signal indicative of voice components contained in the output signal  $\text{OutTfe}(n)$ . In the example system shown, two speech detection signals, a first voice activity detection signal  $\text{VadTfe}(n)$ , e.g., a first VAD flag, and a second voice activity detection signal  $\text{VadFfe}(n)$ , e.g., a second VAD flag, are provided whose generation and characteristics are detailed further below.

[0041] The output signal  $\text{OutTfe}(n)$  is also supplied to a gain control block 403 which additionally receives the speech detection signals and generates the gain control signal  $g(n)$  based on the output signal  $\text{OutTfe}(n)$  and the speech detection signal(s), the first voice activity detection signal  $\text{VadTfe}(n)$  and the second voice activity detection signal  $\text{VadFfe}(n)$ . The gain control block 403 controls the controllable-gain block 401 so that the output signal  $\text{OutTfe}(n)$  is, amplified or attenuated to have a predetermined mean or

maximum or absolute peak signal level as long as no voice components are detected in the output signal  $\text{OutTfe}(n)$ . For example, the output signal  $\text{OutTfe}(n)$  may be blocked, i.e., the output signal  $\text{Out}(n)$  is (almost) zero, as long as no voice components are detected in the output signal  $\text{OutTfe}(n)$ .

5    **[0042]**     The speech leveling system shown in Figure 4 may further include an optional delay block 404 that delays by a certain time (e.g., a certain number of frames or samples) the input signal  $x(n)$  before it is supplied to the controllable-gain block 401 such that the input signal  $x(n)$  supplied to the controllable-gain block 401 is delayed with respect to the input signal  $x(n)$  as it is supplied to the speech detecting block 402 and the gain control  
10   block 403. Situations in which the delay block 404 may be utilized are outlined further below.

**[0043]**     The speech detecting block 402 is further connected to receive, e.g., from a register or memory (both not shown), an SNR threshold  $\text{VadTfeSnr}_{\text{TH}}$ , a threshold  $\text{VadTfe}_{\text{TH}}$ , and, a timer value  $\text{VadTfeTimer}$ , for the first voice activity detection based on  
15   the output signal  $\text{OutTfe}(n)$ , and an SNR threshold  $\text{VadFfeSnr}_{\text{TH}}$ , a threshold  $\text{VadFfe}_{\text{TH}}$ , and, a timer value  $\text{VadFfeTimer}$ , for the second voice activity detection based on an input signal  $\text{InFfe}(n)$  from the far end. For that reason, the input signal  $\text{InFfe}(n)$  is supplied to speech detecting block 402. Further, at least one of a minimum noise threshold  $\text{MinNoise}$ , an initial root mean square (RMS) value  $\text{RmsInit\_dB}$ , and a root mean square threshold  
20    $\text{RMS}_{\text{TH}}$ , may be provided to the speech detecting block 402.

**[0044]**     The gain control block 403 may further receive, e.g., from a register or memory (both not shown), an optional smoothing time parameter  $\tau_{\text{Smooth}}$ , at least one of a maximum gain tuning parameter  $\text{MaxGain}$  and maximum levelling gain tuning parameter  $\text{MaxLevelerGain}$ , and a reference level  $\text{RefLevel}$  or a reference peak level  $\text{RefPeakLevel}$ .  
25   Optionally, a (peak) limiter block 405 may operatively be connected downstream of the controllable-gain block 401 to receive the output signal  $\text{Out}(n)$  and to provide a limited output signal  $\text{Out}_L(n)$ . The (peak) limiter block 405 may further receive signals representing a predetermined limiter gain  $\text{LimGain}$ , a limiter threshold  $\text{LimTh}$ , an attack time  $\text{AttackT}$ , and a release time  $\text{ReleaseT}$ .

[0045] Figure 5 (5A, 5B) is a detailed schematic diagram illustrating an exemplary implementation of the speech leveling system shown in Figure 4. In the speech leveling system as partially shown in Figure 5A, a voice activity detector block 501 may be configured to transform by way of a time-frequency transformation block 502y the input signal in the time domain, which is the output signal  $\text{OutTfe}(n)$ , into a signal in the frequency domain, e.g., a group of discrete spectral components  $Y(\omega)$ . The spectral components  $Y(\omega)$  of the output signal  $\text{OutTfe}(n)$  can be obtained by a variety of ways, including band pass filtering and Fourier transformation. In one approach, a Discrete Fourier Transform (DFT) or Fast Fourier Transform (FFT) may be used to transform sequential blocks of  $N$  points of the input signal. A window function, such as a Hanning window, may be applied, in which case an overlap of a multiplicity of, e.g.,  $N/2$  points can be used. The DFT can be used at each frequency bin in the input signal. Alternatively, the FFT can be used over the whole frequency band occupied by the input signal. The spectrum is stored for each frequency bin within the signal band of the output signal  $\text{Out}(n)$ . Similarly, a time-frequency transformation block 502x applies an FFT with optional windowing (not shown) also to the input signal  $\text{InFfe}(n)$  in the time domain, after being delayed by delay block 500, in order to generate therefrom a signal in the frequency domain, e.g., a group of spectral components  $X(\omega)$ . The spectral components  $X(\omega)$  and  $Y(\omega)$  are herein also referred to just as signals  $X(\omega)$  and  $Y(\omega)$ .

[0046] The signals  $X(\omega)$  may be each optionally smoothed by a group of (e.g.,  $(N/2)$ ) temporal smoothing blocks 503y and the signals  $Y(\omega)$  may be each optionally smoothed by a group of (e.g.,  $(N/2)$ ) temporal smoothing blocks 503x, each block utilizing, e.g., a moving average filter of appropriate length. An (optional) spectral smoothing may account for a plurality of frequency bins, which reduces the spectral resolution. The temporal smoothing blocks 503y and 503x are operatively connected downstream of the time-frequency transformation blocks 502y and 502x, respectively. The groups of temporal smoothing blocks 503y and 503x combine frequency bin values over time for each of signals  $Y(\omega)$  and  $X(\omega)$ , respectively, to reduce the temporal dynamics in the output signal of the time-frequency transformation blocks 502y and 502x.

[0047] The groups of temporal smoothing blocks 503y and 503x output temporal smoothed signals that may still contain impulsive distortions as well as background noise.

Each of the (e.g.,  $N/2$ ) noise estimation blocks 504y and (e.g.,  $N/2$ ) noise estimation blocks 504x is operatively connected downstream of a corresponding one of the multiplicity of temporal smoothing blocks 503y and 503x to smear out impulsive distortions such as speech in the outputs of the temporal smoothing blocks 503y and 503x, and to provide groups of smoothed signals  $S_{Y\_1}(n) \dots S_{Y\_N/2}(n)$  and smoothed signals  $S_{X\_1}(n) \dots S_{X\_N/2}(n)$  that form basis for estimating the respective current background noise. The group of smoothed signals  $S_{Y\_1}(n) \dots S_{Y\_N/2}(n)$  relates to the signals  $Y(\omega)$  and the other group of smoothed signals  $S_{X\_1}(n) \dots S_{X\_N/2}(n)$  relates to the signals  $X(\omega)$ . In order to reduce or avoid smearing of a desired signal such as music or voice signals, non-linear smoothing (not shown) may be employed in the multiplicity of noise estimation blocks 504y and 504x. A group of background noise estimates  $N_{Y\_1}(n) \dots N_{Y\_N/2}(n)$  is determined by one group of the noise estimation blocks 504y and a group of background noise estimates  $N_{X\_1}(n) \dots N_{X\_N/2}(n)$  is determined by the other group of noise estimation blocks 504x.

**[0048]** In a multiplicity of comparator blocks 506y, the one group of smoothed signals  $S_{Y\_1}(n) \dots S_{Y\_N/2}(n)$  is compared to a pre-determined threshold, e.g., an SNR threshold  $Vad_YSnr_{TH}$ , multiplied with the group of background noise estimates  $N_{Y\_1}(n) \dots N_{Y\_N/2}(n)$ , and the other group of smoothed signals  $S_{X\_1}(n) \dots S_{X\_N/2}(n)$  is compared to another pre-determined threshold, e.g., an SNR threshold  $Vad_XSnr_{TH}$ , multiplied with the group of background noise estimates  $N_{X\_1}(n) \dots N_{X\_N/2}(n)$ . The group of smoothed signals  $S_{Y\_1}(n) \dots S_{Y\_N/2}(n)$  and the group of background noise estimates  $N_{Y\_1}(n) \dots N_{Y\_N/2}(n)$  relate to the signals  $Y(\omega)$ , and the other group of smoothed signals  $S_{X\_1}(n) \dots S_{X\_N/2}(n)$  and the other group of background noise estimates  $N_{X\_1}(n) \dots N_{X\_N/2}(n)$  relate to the signals  $X(\omega)$ . If current SNR values determined from smoothed signals  $S_{Y\_1}(n) \dots S_{Y\_N/2}(n)$  and  $S_{X\_1}(n) \dots S_{X\_N/2}(n)$  and background noise estimates  $N_{X\_1}(n) \dots N_{X\_N/2}(n)$  are each greater than corresponding pre-determined thresholds  $Vad_YSnr_{TH}$  and  $Vad_XSnr_{TH}$ , respectively, corresponding flags  $Vad_YSnrFlag$  and  $Vad_XSnrFlag$ , will be set to a first logic value, e.g., '1'. Otherwise, the flags  $Vad_YSnrFlag$  and  $Vad_XSnrFlag$  will be set to a second logic value, e.g., '0'. The flags from each group of the multiplicity of comparator blocks 506 are passed on to a corresponding group summation block 507y and 507x, respectively. The summation block 507y sums up the spectral flags  $Vad_YSnrFlag$  from the multiplicity of comparator blocks 506y and outputs a time varying sum signal  $Y_{VAD}(n)$ . The

summation block 507x sums up the spectral flags  $Vad_{xSnrFlag}$  from the multiplicity of comparator blocks 506x and outputs a time varying sum signal  $X_{VAD}(n)$ .

[0049] The sum signals  $Y_{VAD}(n)$  and  $X_{VAD}(n)$  are passed on to comparator blocks 508y and 508x where the sum signal  $Y_{VAD}(n)$  is compared to a pre-determined sum threshold  $VAD_{Y\_TH}$  and the sum signal  $X_{VAD}(n)$  is compared to a pre-determined sum threshold  $VAD_{X\_TH}$ . If the current value of the sum signal  $Y_{VAD}(n)$  is greater than the pre-determined sum threshold  $VAD_{Y\_TH}$ , a flag  $VAD\_Y(n)$ , which can be considered an updated sum signal  $Y_{VAD}(n)$ , will be set to a first logic state, e.g., '1' (one). Otherwise the flag  $VAD\_Y(n)$  will be set to another logical state, e.g., '0' (zero). If the current value of the sum signal  $X_{VAD}(n)$  is greater than the pre-determined sum threshold  $VAD_{X\_TH}$ , a flag  $VAD\_X(n)$ , which can be considered an updated sum signal  $X_{VAD}(n)$ , will be set to a first logic state, e.g., '1' (one). Otherwise the flag  $VAD\_X(n)$  will be set to another logical state, e.g., '0' (zero).

[0050] Output signals of comparator blocks 508y and 508x, i.e., flags  $VAD\_Y(n)$  and  $VAD\_X(n)$ , are passed on to voice activity/pause detection (VAD/SPD) timer blocks 509y and 509x that detect whether flags  $VAD\_Y(n)$  and  $VAD\_X(n)$  are '0' or not. At a time n, if the flag  $VAD\_Y(n)$  is '0', the VAD timer block 509y sets a count  $T_{VAD\_Y}(n)$  to  $T_{VAD\_Y}(n-1)-1$ , i.e., decrements the previous count  $T_{VAD\_Y}(n-1)$  by one. If not, the count  $T_{VAD\_Y}(n)$  will be reset to an initialization value  $Vad_YTimerInit$ . The output of the VAD timer block 209y, i.e., count  $T_{VAD\_Y}(n)$ , is passed on to a voice activity detection (VAD) block 210y where the count  $T_{VAD\_Y}(n)$  is compared with a pre-determined threshold, e.g., zero. If the current count  $T_{VAD\_Y}(n)$  is greater than the pre-determined threshold (zero), a flag  $VAD_Y(n)$  will be set to '1', otherwise to '0'. Similarly, if the flag  $VAD\_X(n)$  is '0', the VAD timer block 509x sets a count  $T_{VAD\_X}(n)$  to  $T_{VAD\_X}(n-1)-1$ , i.e., decrements the previous count  $T_{VAD\_X}(n-1)$  by one. If not, the count  $T_{VAD\_X}(n)$  will be reset to an initialization value  $Vad_XTimerInit$ . The output of the VAD timer block 209x, i.e., count  $T_{VAD\_X}(n)$ , is passed on to a voice activity detection (VAD) block 210x where the count  $T_{VAD\_X}(n)$  is compared with a pre-determined threshold, e.g., zero. If the current count  $T_{VAD\_X}(n)$  is greater than the pre-determined threshold (zero), a flag  $VAD_X(n)$  will be set to '1', otherwise to '0'.  $VAD_X(n) = 1$  means here that a near end talker is active.  $VAD_X(n)$

= 0 means that the near end talker is not active.  $VAD_Y(n) = 1$  means here that a far end talker is active.  $VAD_Y(n) = 0$  means that the far end talker is not active.

[0051] In the speech leveling system as partially shown in Figure 5B, a gain control block 517 includes a speech root-mean-square (RMS) level determination block 518 which receives the output signal  $OutTfe(n)$ , the flag  $VAD\_Y(n)$ , the flag  $VAD\_X(n)$  and a previous sample of the smoothed RMS signal  $SRMS(n-1)$ . The speech RMS level determination block 518 outputs a signal  $In_{RMS}(n)$ . The signal  $In_{RMS}(n)$  is equal to the output signal  $OutTfe(n)$  if the flag  $VAD\_X(n)$  is '0' and the flag  $VAD\_Y(n)$  is '1', and is set to the value of the a previous sample of the smoothed RMS signal  $SRMS(n-1)$  otherwise. The signal  $In_{RMS}(n)$  is subject to a subsequent maximum absolute value (MaxAbs) determination (calculation) optionally utilizing the initial root mean square (RMS) value  $RmsInit\_dB$ . The value out of the MaxAbs determination is supplied to a smoothing block 519 which provides the smoothed RMS signal  $SRMS(n)$ . The smoothed RMS signal  $SRMS(n)$  is, on the one hand, delayed by one sample to generate the previous sample of the smoothed RMS signal  $SRMS(n-1)$  which is fed back to the speech RMS level determination block 518, and, on the other hand, is supplied to a gain determination block 520 where a gain control signal  $Gain(n)$  is generated according to the ratio between a reference level  $RefLevel$  and the smoothed RMS signal  $SRMS(n)$ .

[0052] In an optional gain limitation block 521, the gain control signal  $Gain(n)$  is compared to a reference value  $MaxVadGain$ . As a consequence, if the gain control signal  $Gain(n)$  is greater than the reference value  $MaxVadGain$ , the output signal of the gain limitation block 521, which is a limited gain control signal  $Gain'(n)$  and which can be considered an updated gain control signal  $Gain(n)$ , is set to the reference voice activity value  $MaxVadGain$ , and otherwise to the gain control signal  $Gain(n)$ . In the present example, the limited gain control signal  $Gain'(n)$  is supplied to a speech pause detection (SPD) gain limitation block 522 which further receives the flags  $VAD_Y(n)$  and  $VAD_X(n)$  as well as a reference speech pause value  $MaxGain$ . If flag  $VAD_X(n)$  is '0' and flag  $VAD_Y(n)$  is '1' at time  $n$ , a final gain control signal  $Gain''(n)$ , which can be considered as updated gain control signal  $Gain'(n)$ , is set to the reference value  $MaxGain$ , and otherwise to the limited gain control signal  $Gain'(n)$ . The final gain control signal  $Gain''(n)$  may optionally be smoothed over time by way of a smoothing filter block (not shown)

before it controls a controllable gain block 523. Gain block 523 is supplied via an optional delay block 524 with the input signal  $\text{OutTfe}(n)$  and which outputs the output signal  $\text{Out}(n)$  which may be processed by an optional peak limiter 225 to provide the limited output signal  $\text{Out}_L(n)$ .

5    **[0053]**    In the system shown in Figure 5, two identical or different tuning sets may be used for the determination of the voice activity detection flags  $\text{VAD}_Y(n)$  and  $\text{VAD}_X(n)$ . However, both are based on spectral SNR values. For example, for the determination of the voice activity detection flag  $\text{VAD}_Y(n)$ , a higher value for the SNR threshold  $\text{Vad}_Y\text{Snr}_{\text{TH}}$  may be used than for the SNR threshold  $\text{Vad}_X\text{Snr}_{\text{TH}}$  for the determination the voice activity detection flag  $\text{VAD}_X(n)$ . This allows, on the one hand, to only set the voice activity detection flag  $\text{VAD}_Y(n)$  accordingly if it is sure that speech is present at the far end, hence the application of a high threshold, and, on the other hand, to set the voice activity detection flag  $\text{VAD}_X(n)$  accordingly only if there is no speech. This may be achieved by setting the respective SNR threshold to a lower value such that even vague speech parts will result in a detected speech activity and, thus, avoid an erroneous setting of the voice activity detection flag  $\text{VAD}_X(n)$ . Further, the numbers of spectral components (lines) that exceed the given signal-to-noise ratio thresholds  $\text{Vad}_Y\text{Snr}_{\text{TH}}$  and  $\text{Vad}_X\text{Snr}_{\text{TH}}$  are determined (flags counted or added) and the consequential numbers  $Y_{\text{VAD}}(n)$  and  $X_{\text{VAD}}(n)$  are compared to thresholds  $\text{VAD}_Y_{\text{TH}}$  and  $\text{VAD}_X_{\text{TH}}$ , respectively. The consequential time varying voice activity detection flag  $\text{VAD}_Y(n)$  will be set to '0', if its initial timer value  $\text{Vad}_Y\text{TimerInit}$  has expired, and the consequential time varying voice activity detection flag  $\text{VAD}_X(n)$  will be set to '0', if its initial timer value  $\text{Vad}_X\text{TimerInit}$  has expired.

25    **[0054]**    If the time varying VAD flag  $\text{VAD}_Y(n)$  is '1', i.e. only in active speech sequences at the far end, and  $\text{VAD}_X(n)$  is "0", i.e. only in speech pause sequences at the near end, the RMS value or, alternatively, the maximum absolute peak value of the output signal  $\text{OutTfe}(n)$  will be determined. Otherwise, the previous sample of the smoothed speech RMS value  $\text{SRMS}(n-1)$  will be passed on to the smoothing filter 519. In order to avoid undesired transient effects, RMS determination may optionally be initialized with a properly chosen initialization value  $\text{RmsInit\_dB}$ . As a smoothing filter, for example, an Infinite Impulse Response (IIR) low-pass filter with a relatively long smoothing time

constant  $\tau$ , may be used. The current gain value  $\text{Gain}(n)$  is determined (calculated) by taking the ratio of the desired reference RMS value of the speech output signal, i.e., reference level  $\text{RefLevel}$  or alternatively reference peak level  $\text{RefPeakLevel}$ , and the output signal of this smoothing filter, i.e.,  $\text{SRMS}(n)$ , so that  $\text{Gain}(n) = \text{RefLevel}/\text{SRMS}(n)$ .

5    **[0055]**     Further, an optional limitation of the maximum of the gain value  $\text{Gain}(n)$  to a value  $\text{MaxVadGain}$  may be applied, in order to avoid successive amplifications of very small input signals, as already described above. Thus, a gain value  $\text{Gain}'(n) = \text{Gain}(n)$  if  $\text{Gain}(n) < \text{GainVadMax}$  and  $\text{Gain}'(n) = \text{MaxVadGain}$  otherwise. Still another limitation of the gain value may optionally be applied, which may act as a level compensator in,  
 10   e.g., hands-free applications and which may be controlled by the time varying voice activity detection flags  $\text{VAD}_Y(n)$  and/or  $\text{VAD}_X(n)$ . Thereby the gain value  $\text{Gain}'(n)$  will be reduced to the given limit  $\text{MaxGain}$  if the voice activity detection flag  $\text{VAD}_X(n)$  assumes '0' and the voice activity detection flag  $\text{VAD}_Y(n)$  assumes '1'. Otherwise  $\text{Gain}''(n) = \text{Gain}'(n)$ .

15   **[0056]**     With the described system it is possible to securely sense speech pauses and as such to apply the needed gating to the gain value  $\text{Gain}(n)$ . A major challenge is to leave this gating stage quick enough once speech (from the near end talker) is active again. If this is not guaranteed, it may happen that the first parts of the on-setting speech may not be amplified as desired and as such cannot be understood, neither by speech recognition  
 20   systems, nor by human listeners (far-end talker). One way to avoid this is to use a quickly reacting voice activity detection and/or speech pause detection. This may be realized by using a non-smoothed, instantaneous RMS value which will be compared to the given threshold as part of the decision. If this is not sufficient, the input signal can optionally be delayed by one or several frames, before the current gain  $G$  is applied to it.

25   **[0057]**     Referring now to Figure 6, which is the system shown in and described in connection with Figure 5 with some modifications. As can be seen particularly from Figure 6A, based on the smoothed signals from the multiplicity of temporal smoothing blocks 503y and 503x, and the estimated quasi stationary background noise signals from the multiplicity of noise estimation blocks 504y and 504x, the variations in the SNR can  
 30   be determined (calculated), e.g., as frequency distribution of SNR values in corresponding subsequent SNR calculation blocks 601y and 601x. Using variations in the SNR, noise

can be differentiated from a desired speech or music signal. For example, a low SNR value may be linked to a variety of noise sources such as an air-conditioner, fan, an open window, or an electrical device such as a computer etc. The SNR may be evaluated in the time domain or in the frequency domain or in the sub-band domain. In groups of comparator blocks 602y and 602x that substitute blocks 506y and 506x shown in Figure 5, output SNR values determined by the groups of SNR determination blocks 601y and 601x are compared to a pre-determined threshold, e.g., SNR thresholds  $VadSnr_{TH}$  and SNR threshold  $SpdSnr_{TH}$ , respectively.

**[0058]** If the determined current SNR values provided by blocks 601y are greater than the pre-determined threshold  $VadSnr_{TH}$ , flags  $VadSnrFlag$  will be set to a first logic value, e.g., '1'. Otherwise, the flags  $VadSnrFlag$  will be set to a second logic value, e.g., '0'. The flags from the multiplicity of comparator blocks 602y are passed on to summation block 507y. The summation block 507y sums up the spectral flags  $VadSnrFlag$  from the multiplicity of comparator blocks 602y and outputs a time varying sum signal  $S_{VAD}(n)$ . The sum signal  $S_{VAD}(n)$  is passed to comparator block 508y where the sum signal  $S_{VAD}(n)$  is compared to a pre-determined sum threshold  $VAD_{TH}$ . If the current value of the sum signal  $S_{VAD}(n)$  is greater than the pre-determined sum threshold  $VAD_{TH}$ , a flag  $S'_{VAD}(n)$ , which can be considered an updated sum signal  $S_{VAD}(n)$ , will be set to a first logic state, e.g., '1'. Otherwise the flag  $S'_{VAD}(n)$  will be set to another logical state, e.g., '0'. The blocks 509y and 510y are operated as shown in Figure 5 with  $VadYTimerInit = VadTimerInit$ ,  $VAD\_Y(n) = S_{VAD}(n)$ ,  $T_{VAD\ Y} = T_{VAD}$  and  $VAD_Y(n) = VAD(n)$ .

**[0059]** If the determined current SNR values provided by blocks 601x are greater than the pre-determined threshold  $SpdSnr_{TH}$ , flags  $SpdSnrFlag$  will be set to a first logic value, e.g., '1'. Otherwise, the flags  $SpdSnrFlag$  will be set to a second logic value, e.g., '0'. The flags from the multiplicity of comparator blocks 602x are passed on to summation block 507x. The summation block 507x sums up the spectral flags  $SpdSnrFlag$  from the multiplicity of comparator blocks 602x and outputs a time varying sum signal  $S_{SPD}(n)$ . The sum signal  $S_{SPD}(n)$  is passed to comparator block 508x where the sum signal  $S_{SPD}(n)$  is compared to a pre-determined sum threshold  $SPD_{TH}$ . If the current value of the sum signal  $S_{SPD}(n)$  is greater than the pre-determined sum threshold  $SPD_{TH}$ , a flag  $S'_{SPD}(n)$ ,

which can be considered an updated sum signal  $S_{SPD}(n)$ , will be set to a first logic state, e.g., '1' (one). Otherwise the flag  $S'_{SPD}(n)$  will be set to another logical state, e.g., '0' (zero). The blocks 509x and 510x are operated as shown in Figure 5 with  $Vad_xTimerInit = SpdTimerInit$ ,  $VAD\_X(n) = S_{SPD}(n)$ ,  $T_{VAD\ X} = T_{SPD}$  and  $VAD_X(n) = SPD(n)$ .

5 [0060] Referring particularly to Figure 6B, block 518 shown in Figure 5 may be modified to substitute block 518 by a block 603 in which the flag  $VAD(n)$  is compared with 1, and if it is equal to 1,  $InRMS(n)$  is set to  $OutTfe(n)$ . Otherwise  $InRMS(n)$  is set to  $SRMS(n-1)$ . As for the rest blocks 518 and 603 are identical. Blocks 519 - 525 are connected downstream of block 603 in the manner shown in Figure 5 in connection with  
10 block 518, wherein  $VAD_X(n) = SPD(n)$  and  $VAD_Y(n) = VAD(n)$ .

[0061] Figure 7 illustrates the system shown in and described in connection with Figure 6 with some modifications, wherein separate frequency signal channels are displayed as one multi-signal channel connection depicted as a bold line. As can be seen particularly from Figure 7A, a delay block 700 providing a temporal signal delay has been  
15 added upstream of the time-frequency transformation block 502y while delay block 500 shown in Figure 5 is omitted. Further, a multi-channel signal path 701 has been added extending from the outputs of SNR calculation blocks 601x to a multiplier block 702. The multi-channel signal path 701 includes a multi-channel divider block 703 that receives the signal-to-noise ratios  $SNR_{FFE}(\omega)$  per frequency of the signal received from the far end  
20  $InFfe(n)$  and divides these signal-to-noise ratios  $SNR_{FFE}(\omega)$  by an SNR threshold  $SNR_{FFE}TH$  to provide a multi-channel masking signal  $Mask_{FFE}(\omega)$ . These masking signals  $Mask_{FFE}(\omega)$  are compared to 1 in a multi-channel comparison block 704 which sets masking signals  $Mask_{FFE}'(\omega)$  to 1 if respective masking signals  $Mask_{FFE}(\omega)$  are greater than 1 and otherwise to the respective masking signals  $Mask_{FFE}(\omega)$ . The masking signals  
25  $Mask_{FFE}'(\omega)$  is then increased by 1 in a multi-channel adder block 505 to provide a multi-channel masking signal  $Mask_{FFE}''(\omega)$ . The multi-channel masking signal  $Mask_{FFE}''(\omega)$  is compared to a minimum masking threshold  $MinMask_{FFE}$  in a multi-channel comparator block 706 which outputs a multi-channel masking signal  $Mask_{FFE}'''(\omega)$ . The multi-channel masking signal  $Mask_{FFE}'''(\omega)$ , which is supplied to the multi-channel multiplier block 702,  
30 is per frequency a fixed value  $MindSpdMask$  if the multi-channel masking signal

$\text{Mask}_{\text{FFE}}(\omega)$  the respective masking signal  $\text{Mask}_{\text{FFE}}(\omega)$  is greater than a minimum masking threshold  $\text{MinMask}_{\text{FFE}}$ , or otherwise the masking signal  $\text{Mask}_{\text{FFE}}(\omega)$ .

[0062] Further, a multi-channel comfort noise generation (e.g., calculation) block 707 has been added to the system. Multi-channel comfort noise generation (e.g., calculation) block 707 receives a multi-channel estimated noise signal  $\text{Noise}_{\text{TFF}}(\omega)$  from the noise estimation blocks 504y and the flags  $\text{VAD}(n)$  from block 510y. Further, comfort noise generation block 707 receives a minimum noise threshold  $\text{MIN}_{\text{CF}}$ , a maximum noise threshold  $\text{MAX}_{\text{CF}}$  and scaling factor  $\text{Scale}_{\text{CF}}$ . The comfort noise generation block 707 generates a noise signal that is added to the output signal Out, e.g., by an adder block 708 inserted between controllable gain block 523 and optional limiter block 525. Comfort noise generation block 707 generates a more pleasant or natural noise to be presented to a listener.

[0063] The description of embodiments has been presented for purposes of illustration and description. Suitable modifications and variations to the embodiments may be performed in light of the above description or may be acquired from practicing the methods. For example, unless otherwise noted, one or more of the described methods may be performed by a suitable device and/or combination of devices. The described methods and associated actions may also be performed in various orders in addition to the order described in this application, in parallel, and/or simultaneously. The described systems are exemplary in nature, and may include additional elements and/or omit elements.

[0064] As used in this application, an element or step recited in the singular and proceeded with the word “a” or “an” should be understood as not excluding plural of said elements or steps, unless such exclusion is stated. Furthermore, references to “one embodiment” or “one example” of the present disclosure are not intended to be interpreted as excluding the existence of additional embodiments that also incorporate the recited features. The terms “first,” “second,” and “third,” etc. are used merely as labels, and are not intended to impose numerical requirements or a particular positional order on their objects.

[0065] The embodiments of the present disclosure generally provide for a plurality of circuits, electrical devices, and/or at least one controller. All references to the circuits, the at least one controller, and other electrical devices and the functionality provided by each, are not intended to be limited to encompassing only what is illustrated and described herein. While particular labels may be assigned to the various circuit(s), controller(s) and other electrical devices disclosed, such labels are not intended to limit the scope of operation for the various circuit(s), controller(s) and other electrical devices. Such circuit(s), controller(s) and other electrical devices may be combined with each other and/or separated in any manner based on the particular type of electrical implementation that is desired.

[0066] It is recognized that any controller or signal processor may include any number of microprocessors, integrated circuits, memory devices (e.g., FLASH, random access memory (RAM), read only memory (ROM), electrically programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM), or other suitable variants thereof) and software which co-act with one another to perform operation(s) disclosed herein. In addition, any controller or signal processor utilizes any one or more microprocessors to execute a computer-program that is embodied in a non-transitory computer readable medium that is programmed to perform any number of the functions as disclosed. Further, any controller as provided herein includes a housing and the various number of microprocessors, integrated circuits, and memory devices ((e.g., FLASH, random access memory (RAM), read only memory (ROM), electrically programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM)) positioned within the housing. The controller(s) as disclosed also include hardware based inputs and outputs for receiving and transmitting data, respectively from and to other hardware based devices as discussed herein.

[0067] While various embodiments of the invention have been described, it will be apparent to those of ordinary skilled in the art that many more embodiments and implementations are possible within the scope of the invention. In particular, the skilled person will recognize the interchangeability of various features from different embodiments. Although these techniques and systems have been disclosed in the context of certain embodiments and examples, it will be understood that these techniques and

systems may be extended beyond the specifically disclosed embodiments to other embodiments and/or uses and obvious modifications thereof.

## CLAIMS:

## 1. A speech signal leveling system comprising:

a controllable-gain block configured to receive an input signal and to generate an output signal by applying a frequency-dependent or frequency-independent controllable gain to the input signal, the gain being dependent on a gain control signal;

a speech detecting block configured to receive the input signal and to generate at least one speech detection signal indicative of voice components contained in the input signal; and

a gain control block operatively coupled with the controllable-gain block and the speech detecting block, the gain control block configured to receive the input signal and the at least one speech detection signal, and to generate the gain control signal based on the input signal and the at least one speech detection signal; wherein the gain control block is further configured to control the controllable-gain block to amplify or attenuate the input signal to have a predetermined mean or maximum or absolute peak signal level as long as voice components are detected in the input signal.

## 2. The system of claim 1, wherein

the speech detecting block is further configured to generate at least one of a voice activity detection signal indicative of voice components contained in the input signal and a speech pause detection signal indicative of speech pauses occurring in the input signal; and

the gain control block is further configured to generate the gain control signal based on at least one of the voice activity detection signal and the speech pause detection signal.

## 3. The system of claim 2, wherein the speech detecting block is further configured to:

determine from the input signal at least two signal-to-noise ratios for at least two frequency bands of the input signal;

compare the at least two signal-to-noise ratios to a signal-to-noise ratio threshold for voice activity detection and to a signal-to-noise ratio threshold for speech pause detection;

provide first signals indicative of whether the at least two signal-to-noise ratios are greater than the signal-to-noise ratio threshold for voice activity detection, and second signals indicative of whether the at least one signal-to-noise ratio is greater than the signal-to-noise ratio threshold for speech pause detection; and

sum up the first signals to provide a first sum signal and to sum up the second signals to provide a second sum signal; wherein

the signal-to-noise ratio threshold for voice activity detection is greater than the signal-to-noise ratio threshold for speech pause detection.

4. The system of claim 3, wherein the speech detecting block is further configured to:

determine a root mean square value or a maximum absolute peak value of the input signal;

compare the root mean square value of the input signal to a root mean square threshold or the maximum absolute peak value to a maximum absolute peak threshold; and

compare at least one of the first sum signal to a voice activity detection threshold and the second sum signal to speech pause detection threshold; wherein the speech detecting block is further configured to at least one of:

provide an updated first sum signal which represents a first logical state if the first sum signal exceeds the voice activity detection threshold and the root mean square value exceeds the root mean square threshold or the maximum absolute peak value exceeds the maximum absolute peak threshold, and provide an updated first sum signal which represents a second logical state otherwise; and

provide an updated second sum signal which represents a first logical state if the second sum signal exceeds the speech pause detection threshold and the root mean square value exceeds the root mean square threshold or the maximum absolute peak value exceeds the maximum absolute peak threshold, and provide an updated second sum signal which represents a second logical state otherwise.

5. The system of claim 1 or 2, wherein the gain control block is further configured to:

determine a mean signal indicative of a root mean square of the input signal, smooth the mean signal with a filter to provide a smoothed mean signal; and take the

ratio between a reference mean level and the smoothed mean signal to provide the gain control signal; or

determine a maximum absolute peak signal indicative of a maximum absolute peak value of the input signal, smooth the maximum absolute peak signal with a filter to provide a smoothed maximum absolute peak signal; and take the ratio between a reference maximum absolute peak level and the smoothed maximum absolute peak signal to provide the gain control signal.

6. The system of any of claims 1 to 5, wherein the gain control block is further configured to at least one of:

control the gain of the controllable-gain block to be at or below a predetermined first gain threshold; and

control the gain of the controllable-gain block to be at or below a second gain threshold that is dependent on the at least one speech detection signal; and

control the gain of the controllable-gain block to block the input signal so that no output signal is provided as long as no voice components are detected in the input signal.

7. The system of any of claims 1 to 6, wherein the system is configured so that a time for transferring the input signal to the controllable-gain block is equal to a time for processing the gain control signal from the input signal in the speech detecting block and the gain control block.

8. A speech signal leveling method comprising:

generating an output signal by applying a frequency-dependent or frequency-independent, controllable gain to an input signal, the gain being dependent on a gain control signal;

generating at least one speech detection signal indicative of voice components contained in the input signal; and

generating the gain control signal based on the input signal and the at least one speech detection signal, controlling the controllable-gain block to amplify or attenuate the input signal to have a predetermined mean or maximum or absolute peak signal level as long as voice components are detected in the input signal.

9. The method of claim 8, further comprising

generating at least one of a voice activity detection signal indicative of voice components contained in the input signal and a speech pause detection signal indicative of speech pauses occurring in the input signal; and

generating the gain control signal based on the voice activity detection signal and the speech pause detection signal.

10. The method of claim 9, wherein generating at least one speech detection signal indicative of voice components contained in the input signal comprises:

determining from the input signal at least two signal-to-noise ratios for at least two frequency bands of the input signal;

comparing the at least two signal-to-noise ratios to a signal-to-noise ratio threshold for voice activity detection and to a signal-to-noise ratio threshold for speech pause detection;

providing first signals indicative of whether the at least two signal-to-noise ratios are greater than the signal-to-noise ratio threshold for voice activity detection, and second signals indicative of whether the at least one signal-to-noise ratio is greater than the signal-to-noise ratio threshold for speech pause detection; and

summing up the first signals to provide a first sum signal and to sum up the second signals to provide a second sum signal; wherein

the signal-to-noise ratio threshold for voice activity detection is greater than the signal-to-noise ratio threshold for speech pause detection.

11. The method of claim 10, generating at least one speech detection signal indicative of voice components contained in the input signal further comprises:

determining a root mean square value or a maximum absolute peak value of the input signal;

comparing the root mean square value of the input signal to a root mean square threshold or the maximum absolute peak value to a maximum absolute peak threshold;

comparing at least one of the first sum signal to a voice activity detection threshold and the second sum signal to speech pause detection threshold; wherein generating at least one speech detection signal further comprises at least one of:

providing an updated first sum signal which represents a first logical state if the first sum signal exceeds the voice activity detection threshold and the root mean square value exceeds the root mean square threshold or the maximum absolute peak value exceeds the maximum absolute peak threshold, and providing an updated first sum signal which represents a second logical state otherwise; and

providing an updated second sum signal which represents a first logical state if the second sum signal exceeds the speech pause detection threshold and the root mean square value exceeds the root mean square threshold or the maximum absolute peak value exceeds the maximum absolute peak threshold, and providing an updated second sum signal which represents a second logical state otherwise.

12. The method of any of claims 8 or 9, wherein generating the gain control signal further comprises:

determining a mean signal indicative of the root mean square of the input signal; smoothing the mean signal with a filter to provide a smoothed mean signal; and taking the ratio between a reference mean level and the smoothed mean signal to provide the gain control signal, or

determining a maximum absolute peak signal indicative of a maximum absolute peak value of the input signal, smoothing the maximum absolute peak signal with a filter to provide a smoothed maximum absolute peak signal; and taking the ratio between a reference maximum absolute peak level and the smoothed maximum absolute peak signal to provide the gain control signal.

13. The method of any of claims 8 to 12, wherein generating the gain control signal comprises at least one of:

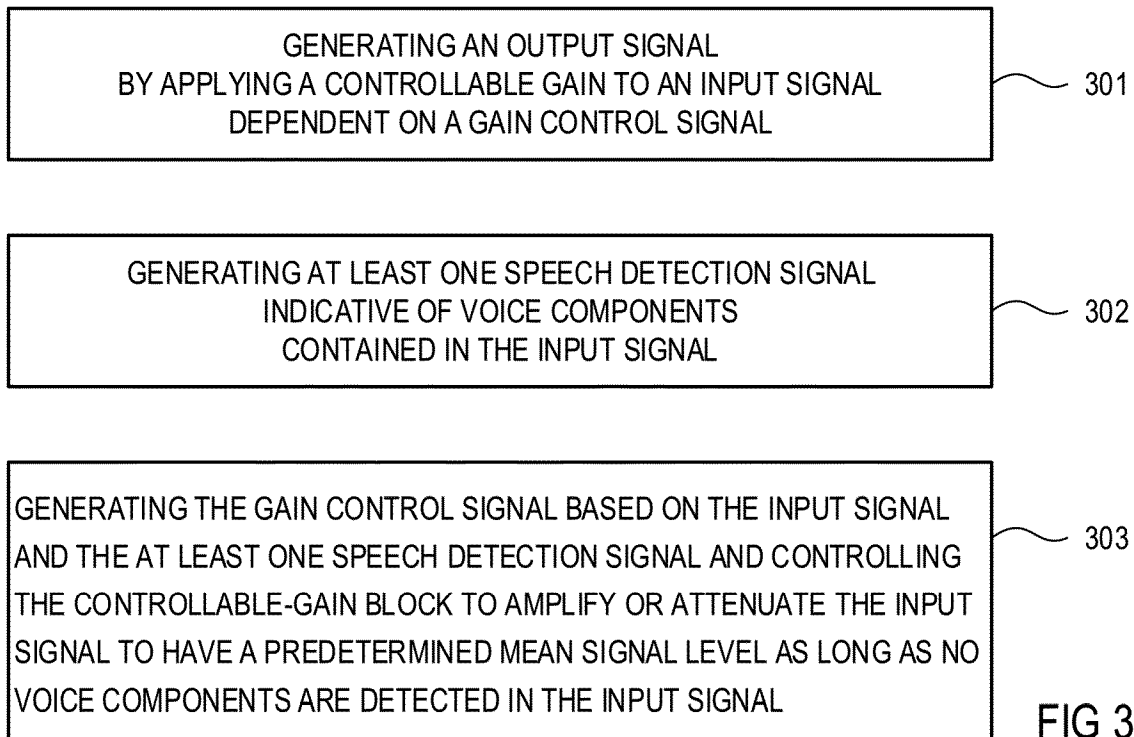
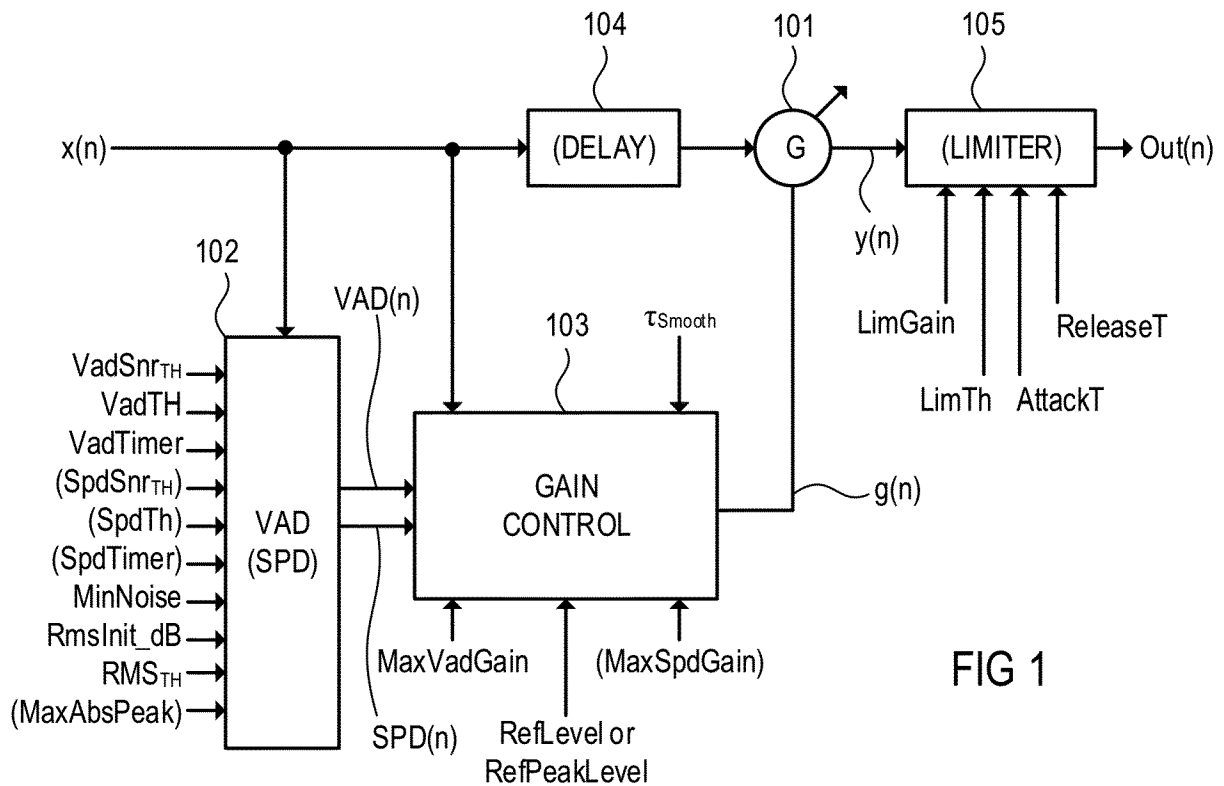
controlling the gain of the controllable-gain block to be at or below a predetermined first gain threshold;

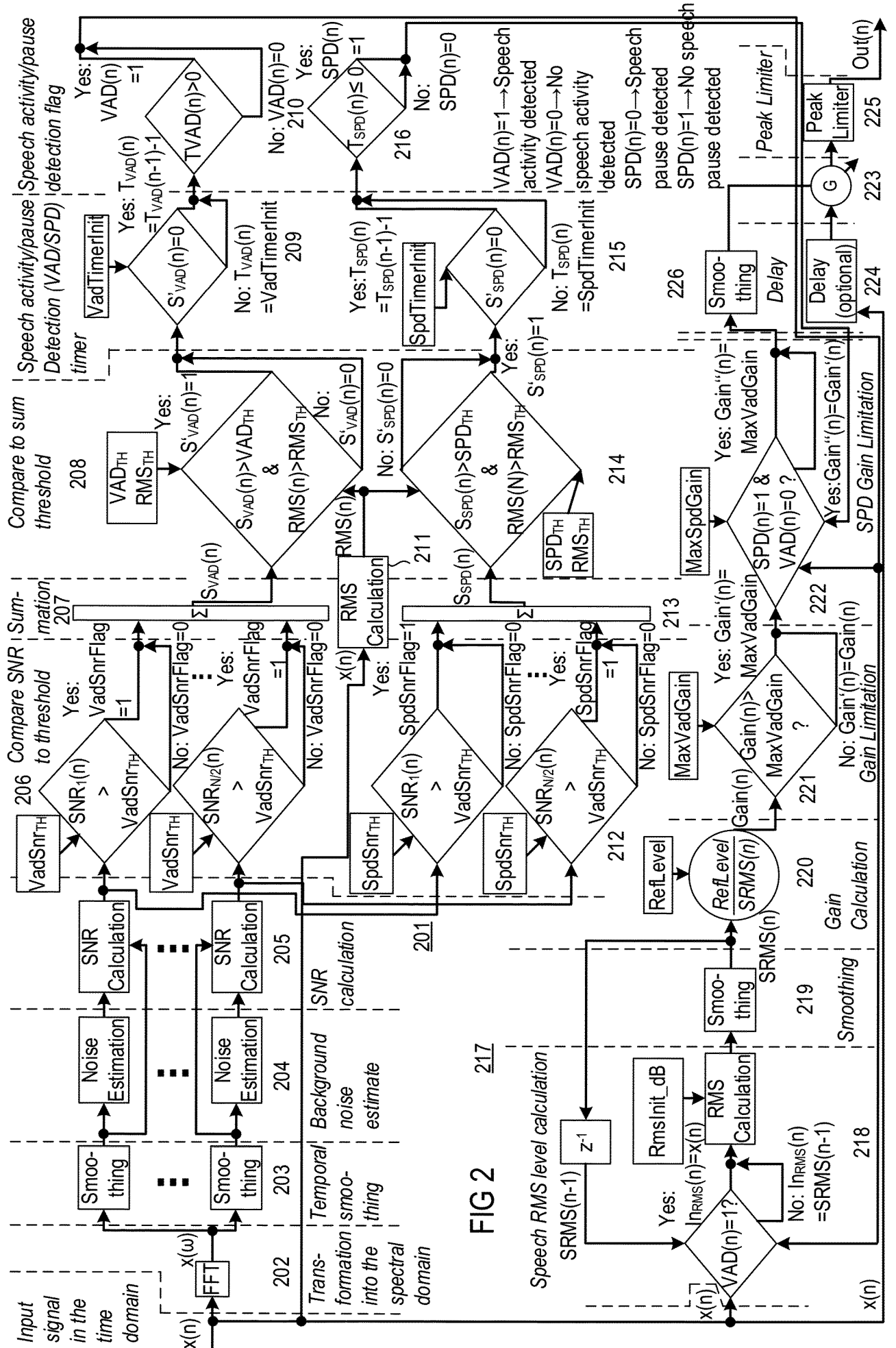
controlling the gain of the controllable-gain block to be at or below a second gain threshold that is dependent on the at least one speech detection signal; and

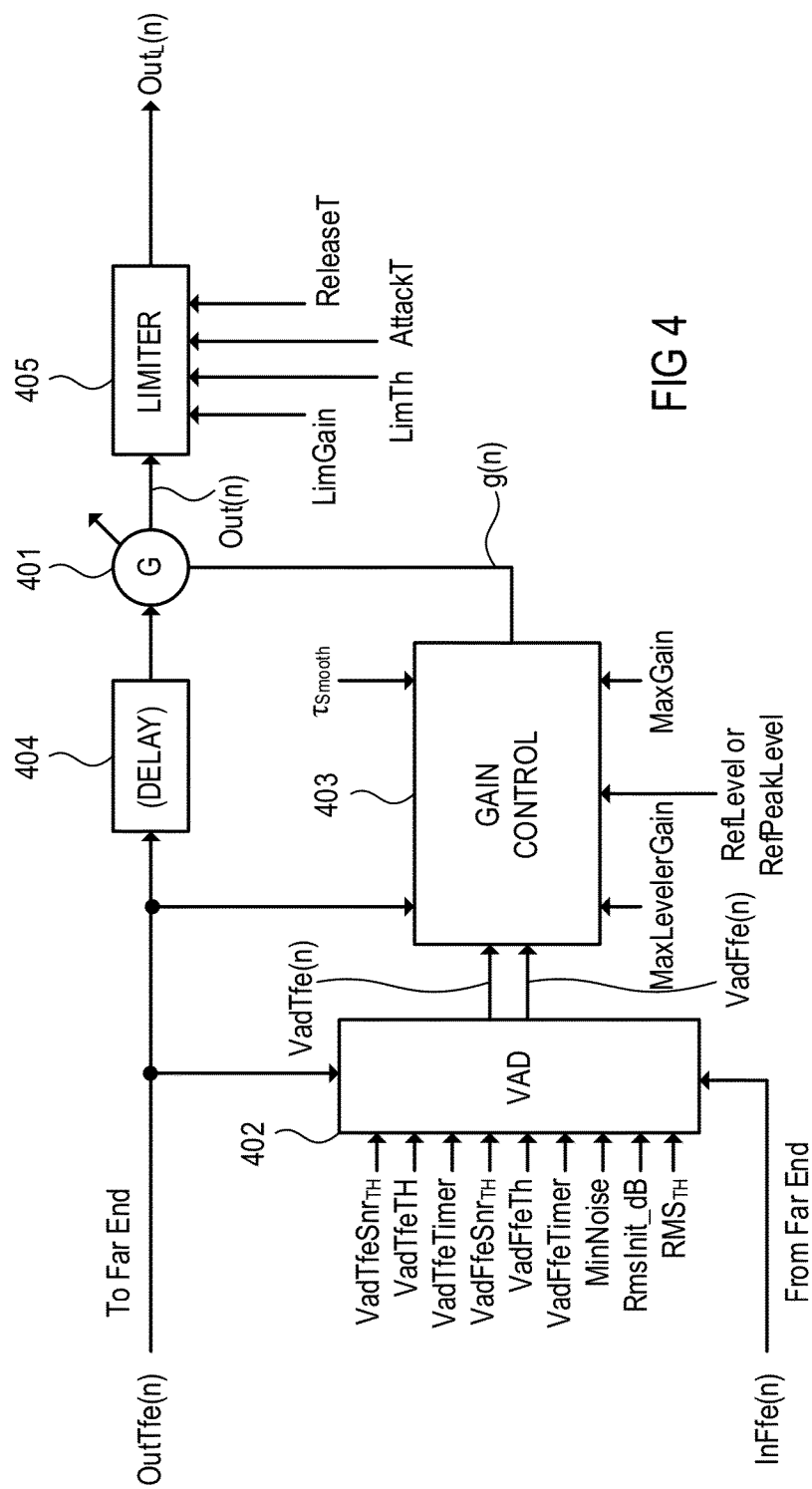
controlling the gain of the controllable-gain block to block the input signal so that no-output signal is provided as long as no voice components are detected in the input signal.

14. The method of any of claims 8 to 13, wherein a time for transferring the input signal to the controllable-gain block is equal to a time for processing the gain control signal from the input signal in the speech detecting block and the gain control block.

15. A computer program, comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method of any of claims 8 to 14.







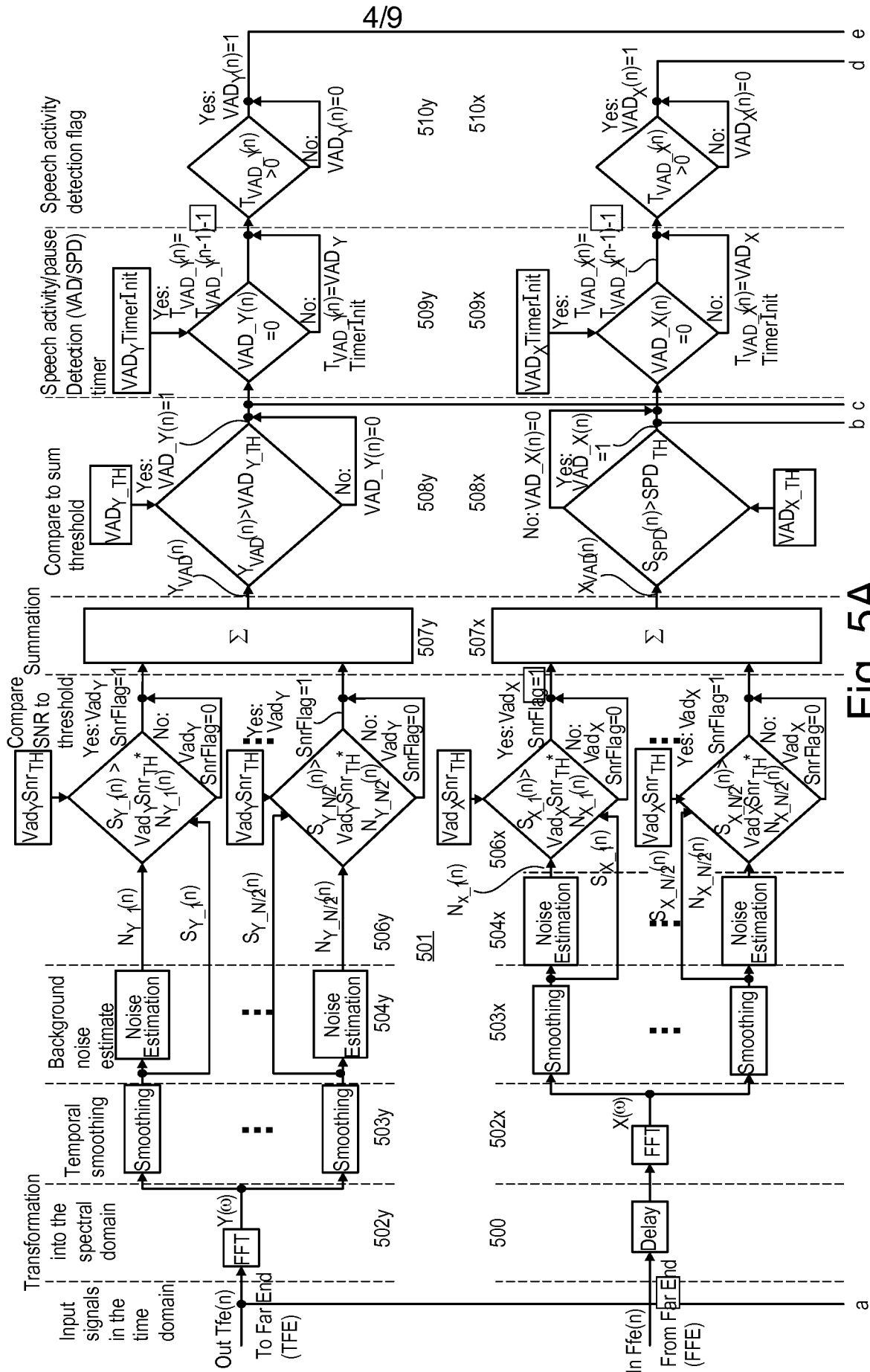


Fig. 5A

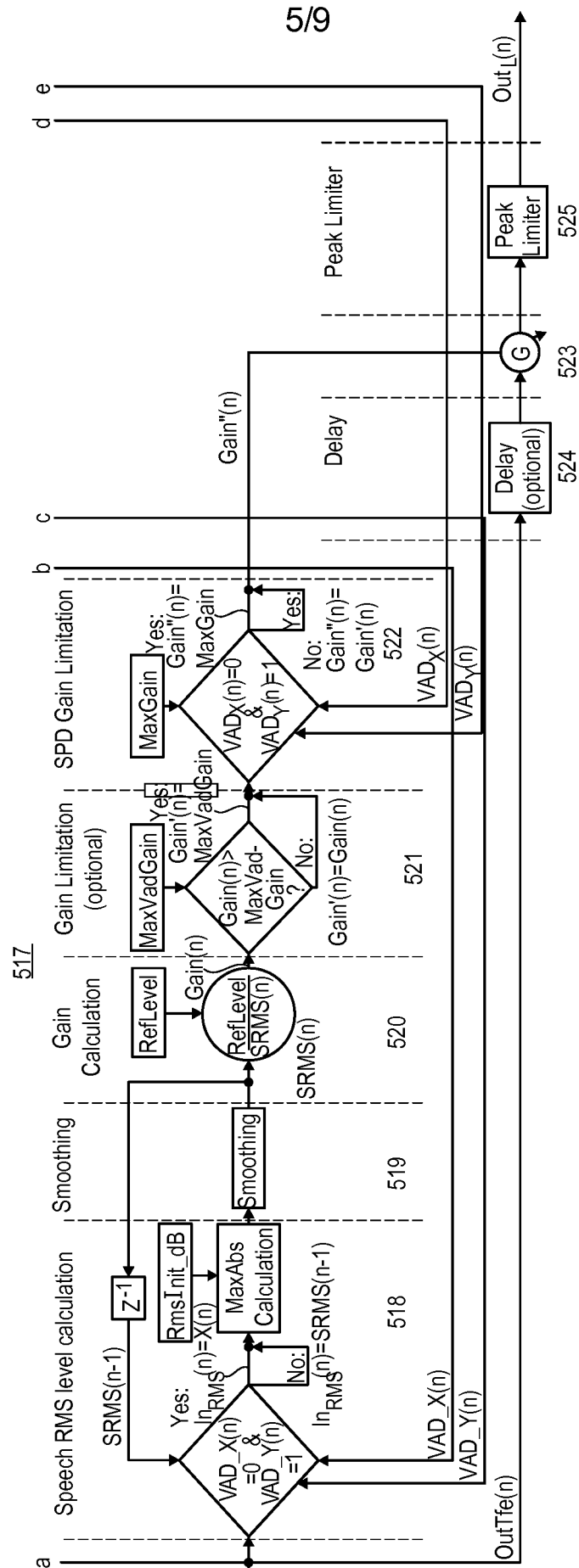


Fig. 5B

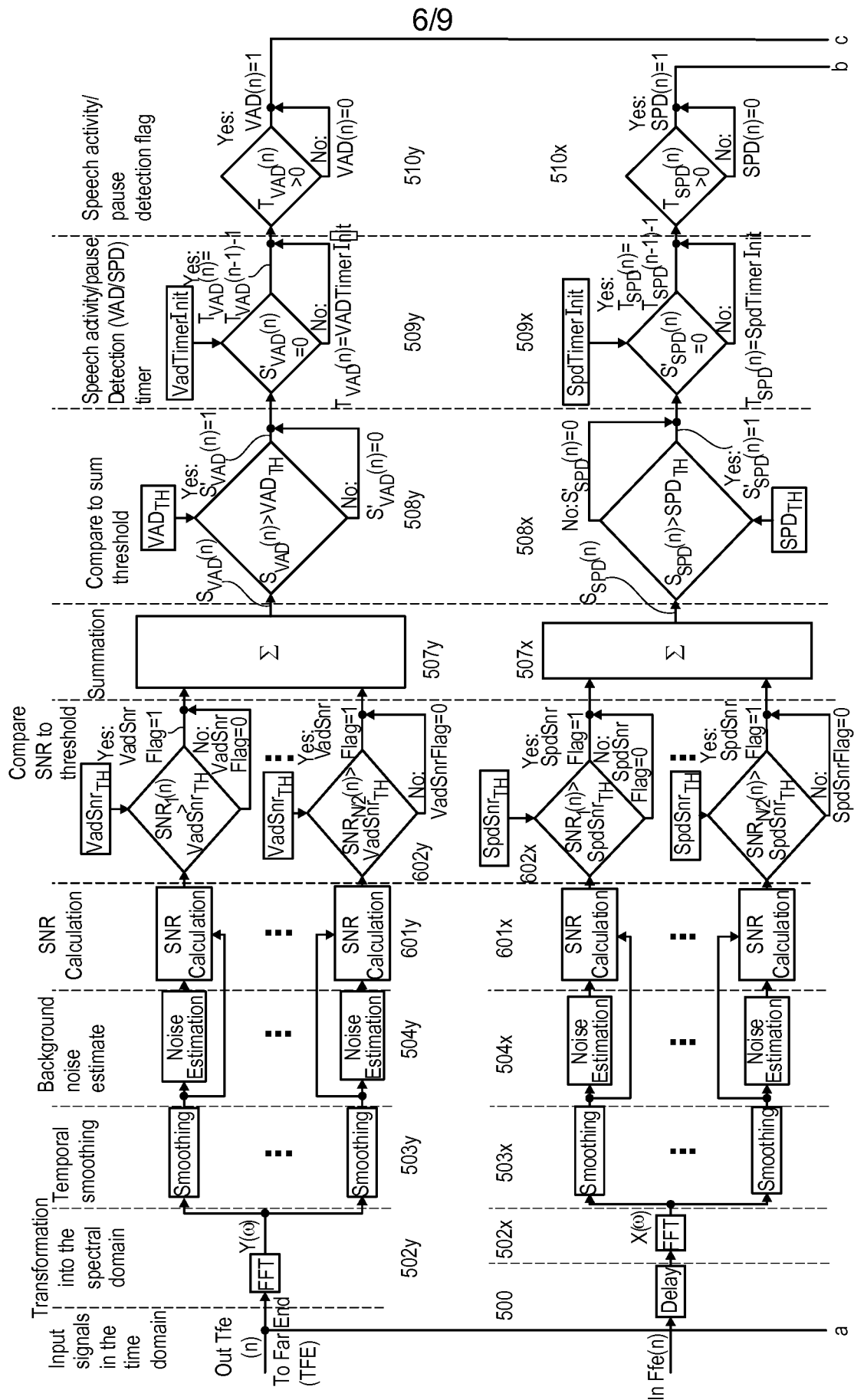


Fig. 6A

7/9

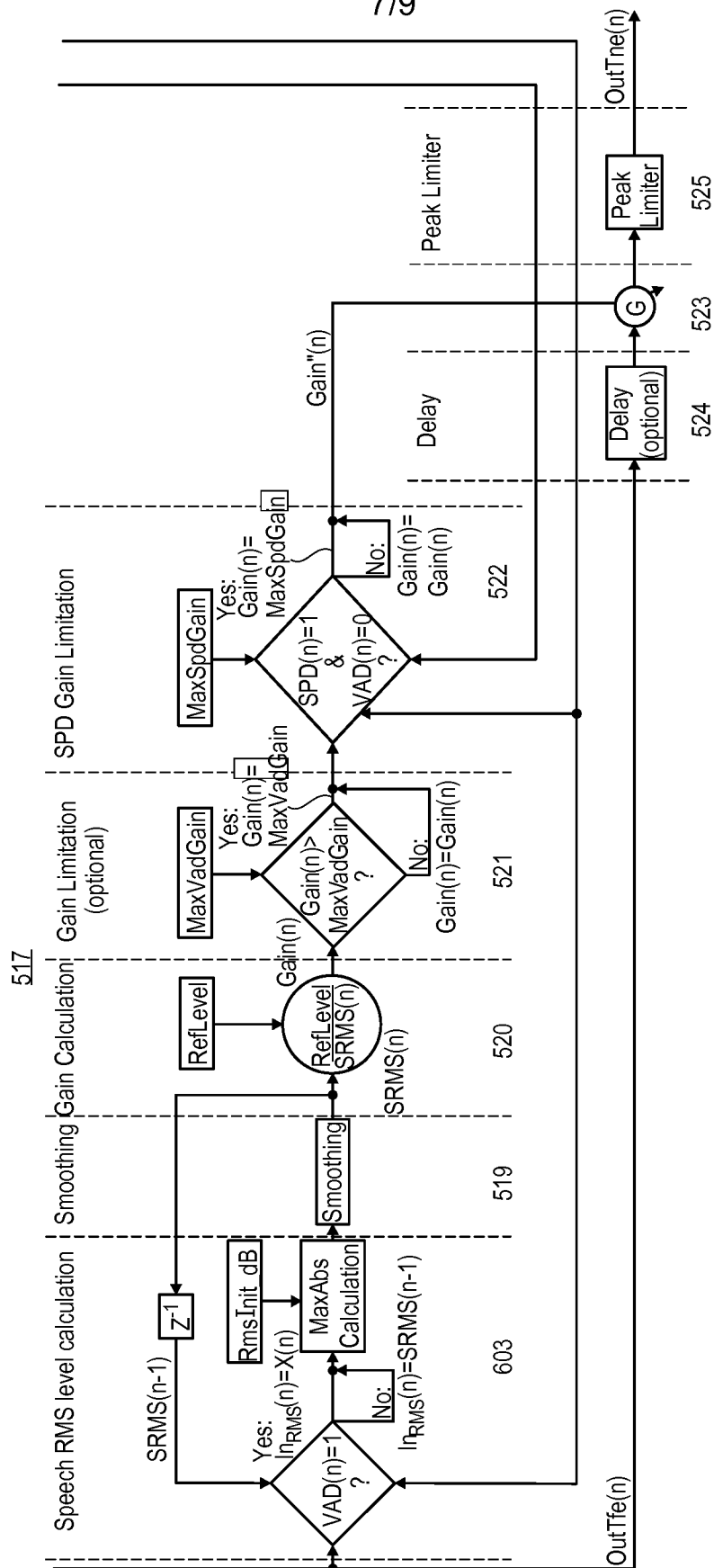


Fig. 6B

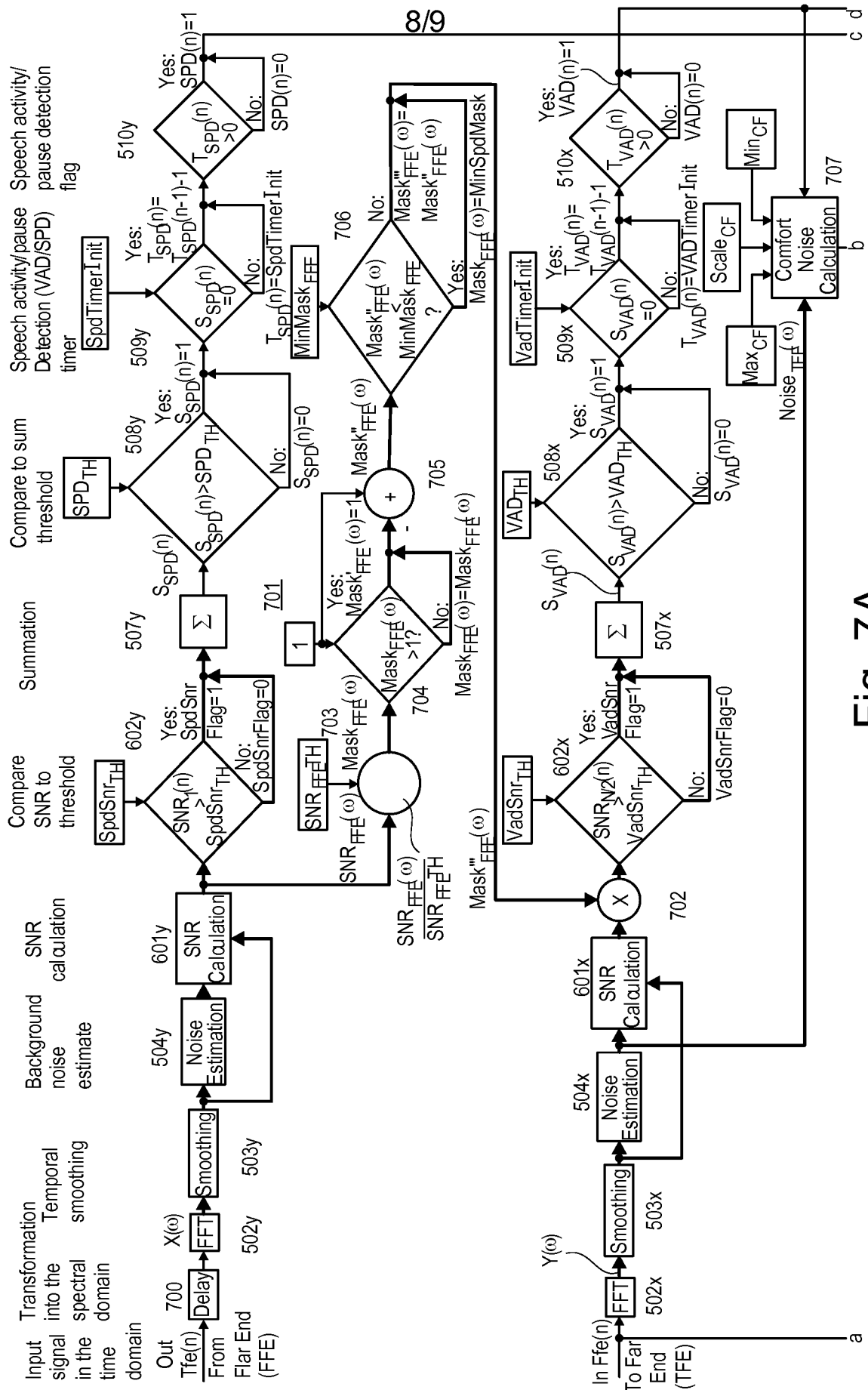


**Fig. 7A**

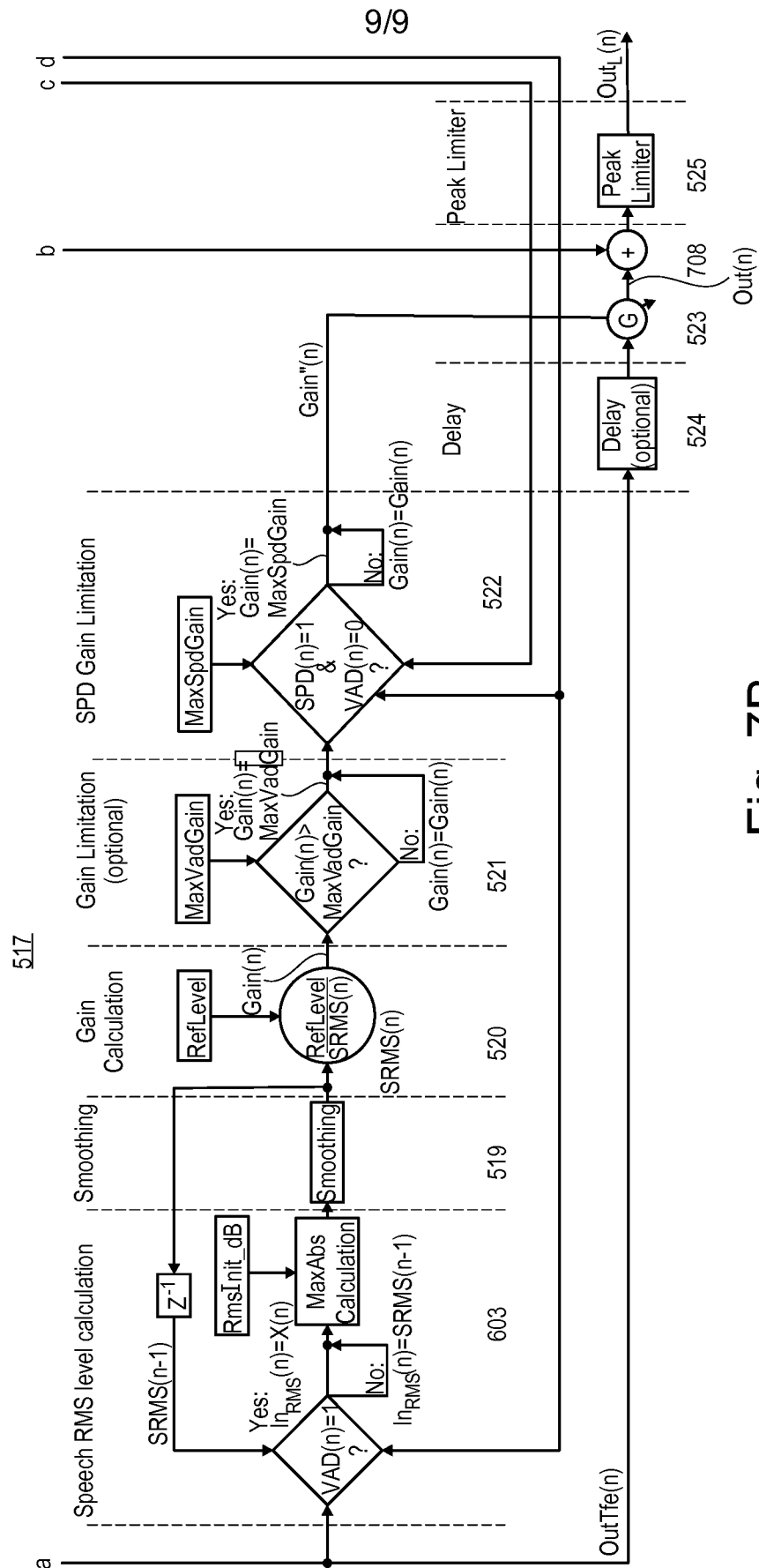


**Fig. 7B**

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2018/069374

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                           |                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>INV. G10L21/0364 H03G3/30 H03G5/16 H03G3/32 H03G3/24<br>H03G7/00 H03G5/00<br>ADD. G10L25/78 G10L25/87                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                           |                                                                                                                               |
| According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                           |                                                                                                                               |
| <b>B. FIELDS SEARCHED</b><br>Minimum documentation searched (classification system followed by classification symbols)<br>H03G G10L                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                           |                                                                                                                               |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                           |                                                                                                                               |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>EPO-Internal, COMPENDEX, INSPEC, WPI Data                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                           |                                                                                                                               |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                           |                                                                                                                               |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Citation of document, with indication, where appropriate, of the relevant passages                        | Relevant to claim No.                                                                                                         |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | US 2010/208918 A1 (NOGUCHI MASAYOSHI [JP])<br>19 August 2010 (2010-08-19)                                 | 1,2,6-9,<br>13-15                                                                                                             |
| Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | figures 12,13<br>paragraphs [0041] - [0043], [0111] -<br>[0116], [0205], [0297] - [0317], [0400]<br>----- | 3-5,<br>10-12                                                                                                                 |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | US 2013/144615 A1 (RAUHALA JUKKA VESA [FI]<br>ET AL) 6 June 2013 (2013-06-06)<br>figures 2,7<br>-----     | 1,8,15                                                                                                                        |
| Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | US 2016/260443 A1 (WANG ZHE [CN])<br>8 September 2016 (2016-09-08)<br>figures 1,2<br>-----                | 3,4,10,<br>11                                                                                                                 |
| Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | US 2013/329912 A1 (KRISHNASWAMY ARVINDH<br>[US] ET AL) 12 December 2013 (2013-12-12)<br>figure 2<br>----- | 5,12                                                                                                                          |
| <div style="display: flex; justify-content: space-between;"> <span><input type="checkbox"/> Further documents are listed in the continuation of Box C.</span> <span><input checked="" type="checkbox"/> See patent family annex.</span> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                           |                                                                                                                               |
| <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p> </div> </div> |                                                                                                           |                                                                                                                               |
| Date of the actual completion of the international search<br><br><div style="text-align: center; font-size: 1.2em;">27 July 2018</div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                           | Date of mailing of the international search report<br><br><div style="text-align: center; font-size: 1.2em;">03/08/2018</div> |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                           | Authorized officer<br><br><div style="text-align: center; font-size: 1.2em;">Chétry, Nicolas</div>                            |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2018/069374

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)                                                                                                                        | Publication<br>date                                                                                          |
|-------------------------------------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| US 2010208918 A1                          | 19-08-2010          | AT 552702 T<br>CN 101807894 A<br>EP 2219371 A2<br>JP 5120288 B2<br>JP 2010192954 A<br>US 2010208918 A1                                            | 15-04-2012<br>18-08-2010<br>18-08-2010<br>16-01-2013<br>02-09-2010<br>19-08-2010                             |
| US 2013144615 A1                          | 06-06-2013          | US 2013144615 A1<br>WO 2011141772 A1                                                                                                              | 06-06-2013<br>17-11-2011                                                                                     |
| US 2016260443 A1                          | 08-09-2016          | CN 102959625 A<br>EP 2619753 A1<br>EP 2743924 A1<br>ES 2489472 T3<br>US 2013304464 A1<br>US 2016260443 A1<br>US 2018061435 A1<br>WO 2012083555 A1 | 06-03-2013<br>31-07-2013<br>18-06-2014<br>02-09-2014<br>14-11-2013<br>08-09-2016<br>01-03-2018<br>28-06-2012 |
| US 2013329912 A1                          | 12-12-2013          | US 2013329912 A1<br>US 2016294343 A1                                                                                                              | 12-12-2013<br>06-10-2016                                                                                     |